

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:20:59 ; Search time 10173.5 Seconds
(without alignments)

13311.631 Million cell updates/sec

Title: US-09-171-553B-3

Perfect score: 8209

Sequence: 1 gtggtgtagcactgtgggcc.....aaaaaaaaaaaaaaaaa 8209

Scoring table: OLIGO-NUC

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_da:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8209	100.0	8209	6 A66553	A66553 Sequence 3
2	7751	94.4	8196	6 A66552	A66552 Sequence 2
3	4833	58.9	8918	14 PEN13816	AJ13816 Porcine e
4	4640	56.5	8849	14 PEN13817	AJ13817 Porcine e
5	4540	55.3	8763	14 PEN293657	AJ293657 Porcine e
6	4532	55.2	8750	14 PEN13818	AJ13818 Porcine e
7	4096	49.9	7808	14 PER117013	Y17013 Porcine end
8	2967	36.1	7333	4 AF038601	AF038601 Sus scrofa
9	2967	36.1	7333	6 AR130474	AR130474 Sequence 1
10	2682	32.7	3320	6 A66551	A66551 Sequence 1
11	2034	24.8	3025	4 AF147808	AF147808 Sus scrofa
12	1971	24.0	1971	14 AF014162	AF014162 Porcine e
13	1499	18.3	1974	6 A66559	A66559 Sequence 9
14	1466	17.9	3482	6 AX002804	AX002804 Sequence
15	1466	17.9	3482	14 PERENV2	Y12239 Porcine end
16	1064	13.0	8918	14 PEN293656	AJ293656 Porcine e
17	876	10.7	927	4 SSU77599	U77599 Sus scrofa
18	848	10.3	4402	6 AX052635	AX052635 Sequence
19	797	9.7	6076	6 AX052636	AX052636 Sequence
20	797	9.7	7362	6 AX052634	AX052634 Sequence
21	746	9.1	6076	6 AR130475	AR130475 Sequence
22	704	8.6	8132	6 AR130475	AR130475 Sequence
23	622	7.6	942	14 MLVR110	X99933 Murine Leuk
24	570	6.9	4918	6 AX052637	AX052637 Sequence
25	557	6.8	8132	4 AF038600	AF038600 Sus scrofa
26	547	6.7	7873	6 AX052638	AX052638 Sequence
27	457	5.6	2462	6 AX002802	AX002802 Sequence
28	457	5.6	2462	14 PERENV1	Y12238 Porcine end
29	446	5.4	1042	14 PERV17012	Y17012 Porcine end
30	437	5.3	2000	14 PEN288588	AJ288588 Porcine e
31	375	4.6	8060	6 AR130473	AR130473 Sequence
32	375	4.6	8132	4 AF038599	AF038599 Sus scrofa
33	367	4.5	1968	14 PEN288589	AJ288589 Porcine e
34	335	4.1	2000	14 PEN288592	AJ288592 Porcine e
35	321	3.9	746	14 PEN298073	AJ298073 Porcine e
36	318	3.9	590	14 PEN298074	AJ298074 Porcine e
37	276	3.4	276	4 SSU54400	AJ005400 Sus scrofa
38	225	2.7	276	4 SSU18745	Y18745 Sus scrofa
39	223	2.7	276	4 SSU54409	AJ005409 Sus scrofa
40	217	2.6	276	4 SSU18748	Y18748 Sus scrofa
41	201	2.4	276	4 SSU18749	Y18749 Sus scrofa
42	189	2.3	276	4 SSU54410	AJ005410 Sus scrofa
43	161	2.0	810	14 AF033259	AF033259 Porcine e
44	158	1.9	522	14 AF033260	AF033260 Porcine e
45	158	1.9	927	4 SSU77600	U77600 Sus scrofa

ALIGNMENTS

RESULT 1

A66553 8209 bp DNA PAT 29-MAR-1999

LOCUS A66553

DEFINITION Sequence 3 from Patent WO9740167.

ACCESSION A66553

VERSION A66553.1 GI:4538106

KEYWORDS

ORGANISM

SOURCE

porcine endogenous retrovirus.

porcine endogenous retrovirus.

Viruses; Retroviral viruses; Retroviridae; Mammalian type C

retroviruses; 1-Mammalian type C virus group.

REFERENCE

1 (bases 1 to 8209)

Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.

TITLE

PORCINE RETROVIRUS

JOURNAL

Patent: WO 9740167-A 3 30-OCT-1997;

Q ONE BIOTECH LTD (GB)

FEATURES

source

Location/Qualifiers

1..8209

/organism="porcine endogenous retrovirus"

/db_xref="taxon:61673"

CDS

588..2162
/codon_start=1
/product="VIRAL CORE PROTEIN"
/protein_id="CAB39347.1"
/db_xref="GI:4538107"
/translation="MGQIVTPTLSLTDHMTENVRSAHNLVQYKKGPWOTFCSEMP
TFEDVWSEGETFENSEIILAVKALIFQTPGSHPOEYILTMDLADPPWPKPMIN
KPKRGRIILALGKRNKSHAEKVEPSSSYLPDGGADLIGTPGSPPTLSSTGCC
TSAPGAVVEGPAAGTGRSRGATPERDEALILPRTYGPAPMGGLQLOLQTPFS
ADLYNMKTNIHPPESDPORLGLIVESLMPHSQHPWDCQOLLOTFTTERENILAE
RKNVGDAGDPFOONETIDMGLFTRGMYNTAEGSESLTKQALVAGLKGASRP
TNLAKEVWQENEPSPVLEFLMEAFRRPTDSEOKASVALAFQSGALDIR
KKIOLGLOAEILRLVREAEKYYRRETEEEKQEKEREEREREROREKNT
KIIAAVEGSSSRERERDFRIRSGPDSGLNRPPLDKQACAYCKEKHMANRNPCK
KNGKPKVLAELEKCD"
<2163..5747
/codon_start=1
/product="REVERSE TRANSCRIPTASE"
/protein_id="CAB39348.1"
/db_xref="GI:4538108"
/translation="GRGSDPLPEPRVTLKVEGQPEVLVDGAEHSVLLQPLKIKE
KSMVMGATGROYPMTTRRTVDLGYGRVHSLVLPCEPVLPLGDLITKMAQIST
EGRPEVSNNKPIIVLTLDDEYRLISDOVADDDISLWEPQOAAEATAGMLA
KQYPOVIOUKASATPSVROYPLSREARGLVPHVQRLIQOILPVOSPMNTPLP
VRKPTNDYRPVODLRVNRKVDIHPVNPYNLISALPERMNTVLDLDAFPCIL
RLHPISQPLFAFEMRDPGRTGQLTWRLPOGFKSPITFEDALHRLANRRIHPO
VTLLQYVDLILLAGATKODLEGTALLESILGVRASAKAOCIRREVTYLGSLR
GGORMLTEARKTIVQIPATTAQVREPLGTAGRCRMLPFPATILADLYLIGEK
GFSMAPEHOKAPAIKALLISAPALAPVYTRFTLYVERKGVAGVLTQILGMR
PVAYLSKIKIDPVAQWPCVCLIAAVALLVKADKLTLYQNTVIAVPALENTVQRP
DRMTNARKHTVOSLLTERVTFAPPAALNPATLLPEETDEVTDHQOLIEETGVR
KDLTDIPLGEVLTWPTDSSVVEGKAGAAVVDGRTIWAASIPETSSOKAEM
ALTOALRLAEKESINITYDSRYAFATVHGAIVKORGLTSAGREIKRELSILE
ATLHPKRLAIICPGHOKADLISRGNDADRAVQAQAVVLLPIETPKAEPERO
YTLDEWQEKIKIDQSEPEGTCTSYCKEILPHKEGLEYVQILHRLHGLHQLQOL
VTPSPHYVRLRGVADSVKRCPCOLVANSPRIPEGRILSHPGHAWEDFTEVK
PAKYGNKYLVLEVDTFSGWVEAVPTKKESTVYAKILIEIPRIGIISDNGPA
FAVQVSQGLAKILIGDKLHCAYRQSSGOVERNMTIKETILKLTTEGINDMALL
PEVLEFRNTPCQGLIPEYELLYGGPPLAIFAFHSAVLDLSLOPLRIRKALEVRO
RAMKOLREAVSGDLOVPHRFQVGDVSVYRHRACNLETRKMKPTLVLLITPTAVKE
GIPTWIIHSHVAPAPPPDSCMAKTEMLPKRLRLVLPYSNNPSGQ"
5620..7590
/codon_start=1
/product="VIRAL ENVELOPE PROTEIN"
/protein_id="CAB39349.1"
/db_xref="GI:4538109"
/translation="MPTLSRHLPTRGCEPKRLRIPSPASIAWFLTLITTPQASK
RLIDSNHPRHLTWLIDDPDGYTVASTGVARGWMLHFLCLMLNPVAKSTP
PRLVSYGFCPCGTGKEKTKGGSGESFCRWSCVTSNDGMKWPISLQDKEFSVA
SGPKYKAMKYLKDKSCSPDLIDYLIKISFTRKTKGSKVKNWELISDFLYGGGA
STLTLRLIETGTEPPVAMGPDVLAEGPALPAPHNLVPOLSLRDPDITOPSPNS
TTGLIPTPTPNRSPGVKTPGQRLFSLIQAFQAINSDPATSNCWCTCSSGPYAE
GMAKRFNNTKEHNOCTGMSRNKLTLEVSQKPCVAVQVLPVRYTHPEVILDED
SENQYLVGYNRMACNTGLPCVYSTSVFNSKPCVAVQVLPVRYTHPEVILDED
YRNPKRPEPYSITLAVMLGTAAGVGTGTAALITGQOLEKGLGELHMAEDLEA
LKEYSNIEBSLTSLSEVLDQNRGLDILLFREGILAALEKCCFVDSGALRDSM
NKLKRELRERRERADQGEFNGFNPSMWTTLISALTGLVLVLLLTITVGPLLINR
FVAFVREVSVOIIVLROQYQGLISQENDL"

CDS

CDS

BASE COUNT 2168 a 2064 c 2154 g 1823 t
ORIGIN
Query Match 100.0%; Score 8209; DB 6; Length 8209;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 8209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaggtacgaactgtgagcccaagcgagcgttggaataaaatccctctgtgttgcac 60
|||||
Db 1 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAAATCCTCTGTGTTGCATC 60
QY 61 aagaccgtctctgtgaatgatttgagggtgtcgcccttcagagcccgagagggagat 120
|||||
Db 61 AAGACCGCTCTCCTCGATGATTGGGGGTGTCCCTCTTCGAGACCCGAGAGGGGGAT 120

QY 121 tttctttacttgaccttcatctgtgtcgttgccgggaatccctgagcaaccctta 180
|||||
Db 121 TGTCTTCTACTGCGCTTTCATTGTGCGCTTGTGGCCGGGAAATCTCGACCACTCTTA 180
QY 181 caccggagacccgaacttgaaggtaaaaggaatccctttggaacgtgtgtgtcgtgcc 240
|||||
Db 181 CACCCGGAACCGACTTGGAGGTAAAGGATCCCTTTGGAACGCTGTGTGTGTGGCC 240
QY 241 ggcgtctgtcttgaaatgtcttcttgagtgatgtgcgtcttcggtttgcaagcttcc 300
|||||
Db 241 GCGCTCTCTTCTTGAATGCTCTTTCGATGATGCGCGCTTTCGCTTTCGACCTGCTCT 300
QY 301 ctaagaacgtaaggaacttgaagactgtatcaacagacgttgaagaagatcaagaagct 360
|||||
Db 301 CTCAGACCGTAAGAGACTGGAGAGCTGTGATCAGACACAGCTGCTAGAGAGATCAGAGCTG 360
QY 361 cagcccttggggagacgcccggagaggttgggagagacgaagacgctgtgtctctact 420
|||||
Db 361 CACCCCTGGGGAGACGCCCGGGAGGTGGGGAGAGCCAGGAGCGCTGTGTCTCTACT 420
QY 421 gtcgtcagagagacgaacttctgtgttgaaacgaagcttccctccgagccgtccg 480
|||||
Db 421 GTGGTACAGAGACCGAATCTCTGTTGTGAACGGAAGGCTTCCCTCGCGCGCTCG 480
QY 481 actctttgcctgtctgttgaagacgagcgagctgcgtgtgtctgtgatactgtgttt 540
|||||
Db 481 ACTCTTTGCTCTGCTGTGGAAGACGCGGAGCGGTCGCTGTGTGATCTGTTGGTTT 540
QY 541 cgtctcgtgtgtcttctgtctgtgtgtgtctgtctgtctgtctgtctgtctgtctgt 600
|||||
Db 541 CGTCTCGTGTGTCTTGT 600
QY 601 tgactaccccccttgcttgactctcgaacttgagcgaagtagatccagggtcata 660
|||||
Db 601 TGACTACCCCCCTTGTTGACTCTCGACCAATTGAGCTGAAGATTAGATCGAGGCTCATA 660
QY 661 atttgcagttcaagttgaagaaggaacttggcagacttctgtgcctctaaatggcaca 720
|||||
Db 661 ATTTGTCAGTTCAAGTTGAAGAAGGACTTGGCAGACTTCTGTGCTCTGAATGGCCAA 720
QY 721 catcagatgtttggtgagcacaagaagggaacttaattcgaataatccctgtgtla 780
|||||
Db 721 CATTCGATGTTGGATGGCCATCAAGAGGACTTAAATTCGAATTAATCTCGCTGTGA 780
QY 781 aggaatcaatcttcagacttgaacgagctctcatccctgacaggaacccatatactta 840
|||||
Db 781 AGGCAATCAATTTTCAAGACTTGACCCGCGCTCTCATCTGATCAGAGACCTATATCTTA 840
QY 841 cgtgcaagatttggcagaagatccctccgcataatgggttaaacatctgtaataaacaa 900
|||||
Db 841 CGTGCAAGATTTTGGCAGAAAGATCCTCGGCATGGTTAAACATGTGCTAATTAACCAA 900
QY 901 gaaagcaaggtccccgaatccctgtctcttgagagaaaacaaacttggccgaaaag 960
|||||
Db 901 GAAAGCAAGTCCCCGAAATCTGCTCTTGTGAAGAAAACAAACATCTGCGCGAAAAG 960
QY 961 tgaagccctctctctgtatctaccccgagatcgagagagcgccgacttggccgaaccc 1020
|||||
Db 961 TCGAGCCCTCTCTCTGATCTAACCCGAGATGAGAGAGCGCCGACTTGCCGGAACCC 1020
QY 1021 caactgttcccccaacccttatcaagcagaggtgtgttgaagggaacttgcctctc 1080
|||||
Db 1021 CAACCTGTTCCCCACCCCTTATTCAGACACAGAGGTGTGTGAGGGAGACTGTGCCCTTC 1080
QY 1081 ctgagatccggtgtgtgagaggaactgtctgtccggagactcgagagcgagccaccc 1140
|||||
Db 1081 CTGAGATCCGGTGTGTGTGAGAGGACTGTCTGCGGAGCTGGAAGCGGACCCACCC 1140
QY 1141 cgaagcgagacagaactcgatatactacgctgcagaccataatggccctccatgagcag 1200
|||||
Db 1141 CGAGCGGAGACAGAGATCGGATATTAACCGCTGCGACCTATGAGCCTCCCATGCGAG 1200

Db 3361 AGGCAGGAGAAAGAACTGATCCAGATACCGGCCCAACACGACCAACAAAGTAGAG 3420
QY 3421 agtlttgggaagcagtgatlttgcaagctgtgacccgggttgcagacttagag 3480
Db 3421 agtlttgggaagcagtgatlttgcaagctgtgacccgggttgcagacttagag 3480
QY 3481 cccgactacccgcttaacaaagaagggggttctctgggtctctgaagcaga 3540
Db 3481 cccgactacccgcttaacaaagaagggggttctctgggtctctgaagcaga 3540
QY 3541 aggaatttgatgctacaaagaagccctgtgagcagcctgtgcttgcctccttagc 3600
Db 3541 aggaatttgatgctacaaagaagccctgtgagcagcctgtgcttgcctccttagc 3600
QY 3601 taactaaacccttaaccccttaagtgtgagcgtgaagggtagcccggaaggtttaa 3660
Db 3601 taactaaacccttaaccccttaagtgtgagcgtgaagggtagcccggaaggtttaa 3660
QY 3661 cccaaacccttgagcactgagggaggaacctgtgtcctacgtgtcaagaagctgtacctg 3720
Db 3661 cccaaacccttgagcactgagggaggaacctgtgtcctacgtgtcaagaagctgtacctg 3720
QY 3721 tagcaggtgtgtgcccgtatgtctgaaagctatcgacagctgtgtgcatctgttgaag 3780
Db 3721 tagcaggtgtgtgcccgtatgtctgaaagctatcgacagctgtgtgcatctgttgaag 3780
QY 3781 acgctgacaaattgacttlttggaagaaatataactgttaataagccccatgtatgaga 3840
Db 3781 acgctgacaaattgacttlttggaagaaatataactgttaataagccccatgtatgaga 3840
QY 3841 acatgtctcgagccccagcagatgtgacaaagccccgagatgacccactataa 3900
Db 3841 acatgtctcgagccccagcagatgtgacaaagccccgagatgacccactataa 3900
QY 3901 gctgtgtctacagagaaggttcaacttgcctcaccagccgctcctaaccctgtcactc 3960
Db 3901 gctgtgtctacagagaaggttcaacttgcctcaccagccgctcctaaccctgtcactc 3960
QY 3961 ttctgcccgaagaaactgtgaacacagtgactgtgacatgagtgacacacacacacacac 4020
Db 3961 ttctgcccgaagaaactgtgaacacagtgactgtgacatgagtgacacacacacacacac 4020
QY 4021 agactgtgtctcgaagaagaccttaccagacatccgctgtgactgtgaggaagtgtcactcgt 4080
Db 4021 agactgtgtctcgaagaagaccttaccagacatccgctgtgactgtgaggaagtgtcactcgt 4080
QY 4081 tcaactgacggaagcagctatgtgtggaaggaaggaatgtgtgtgtgtgtgtgtgtgtgt 4140
Db 4081 tcaactgacggaagcagctatgtgtggaaggaaggaatgtgtgtgtgtgtgtgtgtgtgt 4140
QY 4141 acgggagcccgagaaactgtgagcagcagcctgtcgggaaggaacttcaagcgaagcgt 4200
Db 4141 acgggagcccgagaaactgtgagcagcagcctgtcgggaaggaacttcaagcgaagcgt 4200
QY 4201 agctcatgtgcccctacgcaagccttgcgctgtgcggaaggaatccataaactatata 4260
Db 4201 agctcatgtgcccctacgcaagccttgcgctgtgcggaaggaatccataaactatata 4260
QY 4261 cgaagcagcaggtatgtccttgcgctgtgcgcaagcgtacacggtgtgtgtgtgtgtgtgtgt 4320
Db 4261 cgaagcagcaggtatgtccttgcgctgtgcgcaagcgtacacggtgtgtgtgtgtgtgtgtgt 4320
QY 4321 ggtgtgttactcagcaggaaggaataaagaacaaaggggaatcttcaagcctattag 4380
Db 4321 ggtgtgttactcagcaggaaggaataaagaacaaaggggaatcttcaagcctattag 4380
QY 4381 aagccttactatgtgcaaaagcagctatatacacttctctgacatcagaagcga 4440
Db 4381 aagccttactatgtgcaaaagcagctatatacacttctctgacatcagaagcga 4440
QY 4441 aagatctcatatagaaggaacagagctgtgacgggtgtgcgaagcagcagccag 4500
Db 4441 aagatctcatatagaaggaacagagctgtgacgggtgtgcgaagcagcagccag 4500

QY 4501 ctgttaaccttctgcccataatagaacgcccgaagccccagagaccgaagcagctaca 4560
Db 4501 ctgttaaccttctgcccataatagaacgcccgaagccccagagaccgaagcagctaca 4560
QY 4561 cccatagaagctgtgcaagagataaaaaagatagaccagctctctgaagctcggaggga 4620
Db 4561 cccatagaagctgtgcaagagataaaaaagatagaccagctctctgaagctcggaggga 4620
QY 4621 cctgtctatcctataggggaaggaatcctgtcccaaaaggaaggttagaataatgtcc 4680
Db 4621 cctgtctatcctataggggaaggaatcctgtcccaaaaggaaggttagaataatgtcc 4680
QY 4681 aacagatactgtctcaaccacactgaagacctaacaacctgtgagcagctgtgtgacagacat 4740
Db 4681 aacagatactgtctcaaccacactgaagacctaacaacctgtgagcagctgtgtgacagacat 4740
QY 4741 ccccttactatgtctgaagctacagagagtgagctgtgactgtgtgtcaacatltgtgtc 4800
Db 4741 ccccttactatgtctgaagctacagagagtgagctgtgactgtgtgtcaacatltgtgtc 4800
QY 4801 cctgtcagagctgtgtatgtctaatccttccagaatacctccaggaaggaaggaagga 4860
Db 4801 cctgtcagagctgtgtatgtctaatccttccagaatacctccaggaaggaaggaagga 4860
QY 4861 gccaccagcgtcactcagtggaagtgagcttcaactgaagtaagccggttaataagga 4920
Db 4861 gccaccagcgtcactcagtggaagtgagcttcaactgaagtaagccggttaataagga 4920
QY 4921 acaaatatctatgtgttltgtgagacactlttcagaatgtgtgaaggtctactacta 4980
Db 4921 acaaatatctatgtgttltgtgagacactlttcagaatgtgtgaaggtctactacta 4980
QY 4981 agaaagagacttcaaccgtgtgtgtgtaagaataatctggaaggaatlttccaaagtgtg 5040
Db 4981 agaaagagacttcaaccgtgtgtgtgtaagaataatctggaaggaatlttccaaagtgtg 5040
QY 5041 gaatacctaaggaatgaaggaatgaagcaatgtgtcagcttctgtccaggtgaagtcag 5100
Db 5041 gaatacctaaggaatgaaggaatgaagcaatgtgtcagcttctgtccaggtgaagtcag 5100
QY 5101 gactgtgcccgaagatgt 5160
Db 5101 gactgtgcccgaagatgt 5160
QY 5161 caggaacaggtagaaggaatgaagcaatgaagcaatgaagcaatgaagcaatgaagcaatga 5220
Db 5161 caggaacaggtagaaggaatgaagcaatgaagcaatgaagcaatgaagcaatgaagcaatga 5220
QY 5221 agactgtgacatgaatgt 5280
Db 5221 agactgtgacatgaatgt 5280
QY 5281 cccctgtgacaggt 5340
Db 5281 cccctgtgacaggt 5340
QY 5341 cagaatltgctltgcacataatgt 5400
Db 5341 cagaatltgctltgcacataatgt 5400
QY 5401 tcaagcgtcgt 5460
Db 5401 tcaagcgtcgt 5460
QY 5461 gaggagactgtgaaatgt 5520
Db 5461 gaggagactgtgaaatgt 5520
QY 5521 accgtgtcagaaacctcagagactcgt 5580
Db 5521 accgtgtcagaaacctcagagactcgt 5580

QY 5581 caacgctgtgaaagtcgaagaatcccccctgcatccatcccaagcttaacccg 5640
|||||
Db 5581 CAAGGCTGTGAAGTCGAGGAATCCCCCTGGATCCATGATCCCAAGTTAAACCCG 5640
QY 5641 cgcacaccccgagctcgggttgagagccgaaaaagactgaaatccccccttaagctcgc 5700
|||||
Db 5641 CGCACCTCCCGAGCTCGGGGTGAGAGCCGAAAAAGACTGAGATCCCTTAAGCTTCGCC 5700
QY 5701 tccatcgctgtgtccttactcttaacaataaccccccagccagtgatgaagccttata 5760
|||||
Db 5701 TCCATCCCTGGTTCCTTACTTAAACATATCCCGAGCCAGTATGAACCCCTTATA 5760
QY 5761 gaacgctcgaacccccaatagaccttaacccttaacctgtgtatcttaacccttatac 5820
|||||
Db 5761 GACAGCTCGAACCCTTAACCTTTATCCCTTAACCTGGTATTTATGACCTGATAGC 5820
QY 5821 ggtgcatctgtaaatagcactcgaggtgtgtctccttaagagcaacctgtgtcctgaag 5880
|||||
Db 5821 GGTGTCACTGTAATAGCAGCTCGAGGTGTGCTCTAGAGGACACCTGTGGCTGAACCTG 5880
QY 5881 cattctgcctcgattgattaaaccccgctgttaaaaagcacctcccaacctagtcct 5940
|||||
Db 5881 CATTTCGCTCCGATTGATTAAACCCGCTTTAAAGCACACCTCCAACTAGTCCTG 5940
QY 5941 agttaaggtctctatgtctgccagagcagagaaagaaatctgttgaggtctcag 6000
|||||
Db 5941 AGTTATGGGTCTATTTGCTGCCCCAGCAGACAGAAAGAAATATCTGTGGGGTTCCTGCG 6000
QY 6001 gaatcctctgtagaagaatgagctgtcgtcaacctccaaagatgagactgagaaatggccg 6060
|||||
Db 6001 GAATCTCTGTAGAGATGAGCTGCGTCACCTCCAACATGAGACTGAAATGCGCC 6060
QY 6061 atctctctcagagccgggttaaaatctcctctgttcaatctccgcccgggcaagtcgaa 6120
|||||
Db 6061 ATCTCTCTCCAGGACCGGTAATAATTCCTTTGTCATTTCCGGCCGGCAAGTAAATA 6120
QY 6121 atgtagaactataaagaatagagctgtcctcccaatcaagactatagatcttaaaagata 6180
|||||
Db 6121 ATGATGAACATATATTAAGATAGAGCTGCTCCCATCAGACTTAGTTATTTCTAAATA 6180
QY 6181 agttcactgaaagaaacagagaaatlatcaaaagtgaataatgtgtatgagctgag 6240
|||||
Db 6181 AGTTCACTGAAGGAAACAGGAAATATTTCAAAAGTGAATATGATGAGCTGGCG 6240
QY 6241 aatagtttttatatatgtcgggggagccgggtcccaacttaacatctgccttagata 6300
|||||
Db 6241 AATAGTTTTTATATATGCGGGGAGCAGAGGTCCTTTAACCTATTCGCTTAGGATA 6300
QY 6301 gagaagggagacagacccctgtgtgcaatgagcccgataaagactgagctgaaacaggg 6360
|||||
Db 6301 GAGAGGGGAGAGAACCCCTGTGGCAATGGGACCCGATTAAGTACTGGCTGAACAGGG 6360
QY 6361 ccccgagccctgagcaacgcatactgtccggtgccccaaatlaacctgctgtggcct 6420
|||||
Db 6361 CCCCCGGGCTGAGAGCCACCGCATATACCTGCGGTGCCCAATTAACCTGCTGGCGCT 6420
QY 6421 gacataaagcccgccgaagcaagcaatgacactgtatgttctctacaacagccataga 6480
|||||
Db 6421 GACATTAACACAGCGCGCTAGCAGACATGACACTGATTTGATTCCTTACCAACAGCCCTAGA 6480
QY 6481 aactcccaaggtgtctcgttlaagaagagacagagactctcaagctccatcccaagggact 6540
|||||
Db 6481 AACTCCCAAGGTGTCTCGTTAAGACAGAGACAGACTCTCACTGATCCAGAGGACT 6540
QY 6541 ttccaagccatcaactcccaagccatgagccatctctctgtgtgtgtctatcc 6600
|||||
Db 6541 TTCCAAAGCCATCACTCCACAGCCAGCTGATGCCACTTCTTCTGTGGCTTTGCTATCC 6600
QY 6601 tcaaggtcctctatctatgaagagatgagcttaaaagaaataatcaatgtgacaaagag 6660
|||||
Db 6601 TCAGGGCTCTCTATTTATGAGGGGATGAGCTTAAGAAAGAAATTCATGTGATCAACAGAG 6660
QY 6661 catagaataatcaatgtacatgtgggtgccgaaataagcttaacctcaactgaagttccgg 6720
|||||

Db 6661 CATAGAAATCAATGATGAGGGGTCCGAAATAGCTTACCTACGTAAGATTCCGG 6720
QY 6721 aagggacatgtataggaagaagctccccatcccaacacacttgctatagactgtg 6780
|||||
Db 6721 AAGGGACATGTGATAGGAAAGCTCCCCATCCACACACCTTTGCTATATGATCTGTG 6780
QY 6781 gcttaagcaagccctcagaataatcagatattagactgtgtatataaagatgtgtgagca 6840
|||||
Db 6781 GTTTATGACACAGCCCTCAGAAATCAGTATTTAGTACCTGTTATACAGGTTGGGCA 6840
QY 6841 tgcataactgtgttaacccccgtgtgttccacctagcttcaacccaatccaagaatttc 6900
|||||
Db 6841 TGCATATGCTGCTTAACCCCTGTGTTCACACTAGCTTCAACCAATTCAGAAATTC 6900
QY 6901 tglgtcatgttccaaatgttcccccaggtgtactcaacctctgaggaagtgtctctgat 6960
|||||
Db 6901 TGTGTATGTGCTCAAAATCTGCTCCCGAGTGTACTACATCTCGAGAGTGTCTTGTAT 6960
QY 6961 gaataactatcgtatataccgacccaagagacccgtatcccttaacctagctgta 7020
|||||
Db 6961 GATATGACTATGCTATACCGACCAAAAGAGAACCCGTATCCCTTAACCTAGCTGTA 7020
QY 7021 atgtcgtactaggaagccgtgtgtgaggaagagagagagagagagagagagagagag 7080
|||||
Db 7021 ATGCTCGATTTAGGAGCGCCCTGTGGCTAGGACAGGAGCAGCTGCTGATCACAGGA 7080
QY 7081 ccaagcaactagagaagaagactgtgtgactatgtgcagccatgacagaatctccga 7140
|||||
Db 7081 CCACAGACTAGAGAAAGAGACTGTGTGACTACATGCGGCTATGACAGAAATCTCCGA 7140
QY 7141 gcttaagagagctgtgttagcaacctagagaagctccctgactctgtgtcgaagtgt 7200
|||||
Db 7141 GCCTTAAGAGAGCTGTGTAGCAACCTAGAAAGTCCCTGACTTCTTGTGAAGGTT 7200
QY 7201 ctacagaaccggagagatagatctgtcgttcttaaggaaggtgtgtatgtgcagcc 7260
|||||
Db 7201 CTACAGAACCGGAGGGGATTTAGATGCTGTCTTAAAGAGAGGTGGTTATGTGAGCC 7260
QY 7261 ttaagaagaagatgtgtcttctctatgtatagactcaagagccatcaagatctcaaac 7320
|||||
Db 7261 TTAAAGAAAGATGTGTCTTCTATGTAGATCACTAGAGGACATCAGAACTCAGAAAC 7320
QY 7321 aagcttagaagaagatlaagagagcgtcgaaggaagaagaagagctgacaaaggtgtgt 7380
|||||
Db 7321 AACCTTAGAAAGAAAGTTAGAGAGCGCTGCAAGGAAAGAGAGCTTACAGAGGGTGT 7380
QY 7381 gaagatgtgtcaacaggtctcctgtgatagaacacccctgtcttctgtcagcggggccc 7440
|||||
Db 7381 GAAGGATGTGTTCAACAGGTCCTGTGATGACACCCCTCTTCTGTGACGGGGCC 7440
QY 7441 cttagtactgtctcctgttacttaagttgtggccttgcttaataatagattgtgtgtcc 7500
|||||
Db 7441 CTAGTACTGTCTCTCTCTGTTACTTACAGTTGGGCTTCTCTTAATTAATGTTGTGTG 7500
QY 7501 ttgttagagaagatgtgtgagtcagtcagatcaatgagttatgagcaacagtaacaaagc 7560
|||||
Db 7501 TTTGTTAGAGAGAGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 7560
QY 7561 ctctcagccaaagagaagaactgagcctctagccttccagttctaaatagaaacttaata 7620
|||||
Db 7561 CTCTCGAGCCAGAGAGAAATGACCTCTTACCTCTTCCAGTTCTTAATTAATTAATTAAT 7620
QY 7621 caagaagaagatgtgggaatgaagaatgaagaatgaagaatgaagaatgaagaatgaaga 7680
|||||
Db 7621 CAAGACAGAGAGTGGGGAATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 7680
QY 7681 aagttataaagaacttaaatgtccccgaatatacagaccctgtgtgtcgaagtaata 7740
|||||
Db 7681 AAGTTATAAAGAGCTTAATGCTTAATGCTTAATGCTTAATGCTTAATGCTTAATGCTTAAT 7740
QY 7741 ggtagaaggtcacactctctatgttccaggggctctctcctgtgagtaagataaac 7800
|||||

QY 1333 accagcctacttgga tga tgt lcaacagctgc lgcagacactlccacacccgagagc 1392
|||||
Db 1321 ACCAGCCTACTTGGATGATGTGTCACACAGCTCTCGAGACACTTTCACAAACCGAGAGC 1380
QY 1393 gagaagaattctgttagagcttagaaaaatgtccctggggccagagggagaccacagc 1452
|||||
Db 1381 GAGAGAGAAATCTGTAGAGGCTAGAAAAATGTTCTGGGGCCACGCGGACCCACGC 1440
QY 1453 agttgcaaaatagatgtgacatgga tttcccttgactgcgcgcggttggaactaca 1512
|||||
Db 1441 AGTTGCCAAATGAGATTGACATGGATTCCCTTGACTGCCCCGGTTGGGACTACACA 1500
QY 1513 cggctgaagg taaggagagcttgaaaaatcatalcgcaagcctcgtg lgcgggtccgg 1572
|||||
Db 1501 CGGCTGAAGTAGGAGAGCTTGAAATATCATGCGAGGCTCTGGGGGGGCTCTCCGG 1560
QY 1573 ggcctcaaga cggccactaatltggctaaagttaagaaagtgaatgagagaccgaagc 1632
|||||
Db 1561 GCGCCTCAAGCGGCCACATAATTTGGCTAAGGTAAAGAGGTGATGCGAGGACCGAACG 1620
QY 1633 aaactccctcgtatcttcttgagaggtlca tggaaagcctlcaagcggtlccaccccttg 1692
|||||
Db 1621 AACCTCCCTCGTATTCTTGAGAGGCTCATGGAAGCTTCAGGCGTTACCCCTTTTG 1680
QY 1693 atcctacccagagcccaagaagcctcag tggccctggcctcacttgaggcagtggtc 1752
|||||
Db 1681 ATCCACTCAGAGGCCCAAGAACCTCAGTGGCCCTGGGCTTCATTGGGCACTCGGCTC 1740
QY 1733 tggatataaga aaaaactlcaagagacttgaagaggtlcaagagagctgaatctcgtgac 1812
|||||
Db 1741 TGGATATCGAGAAAGAACTTCAGAGACTGGAAGGTTACAGAGGCTGAGATTACGTGATC 1800
QY 1813 tagtgaagagcagagaaaggttgtatlcagaaagggagacagaaagagagaaagacaga 1872
|||||
Db 1801 TAGTGAGAGAGCAGAGAAAGGTGATTACAAAGCAGACAGAAAGAGAGAAAGAACAGA 1860
QY 1873 gaaaaaagaagagagagaaagaaagagagagagagagagagagagagagagagagagag 1932
|||||
Db 1861 GAAAAAGAAAGAGAGAAAGAAAGAGAGAGAAAGAGAAAGAGAGAGAGAGAGAGAGAG 1920
QY 1933 tgaactaagatcttgccagctggtltgaagaaagcagcagagagagagagagagagat 1992
|||||
Db 1921 TGACTAAGATCTTGGCCGACAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAT 1980
QY 1993 ttaagaaaattagtgccagccctagacag tcaaggaacctgggcaataagagccactcg 2052
|||||
Db 1981 TTAGGAAATTAGTGTCAGGCTTACAGAGTCAAGGAACTGGGCAATGAGACCCACTCG 2040
QY 2053 acaagagcagatgttcgtatgtltaagaaaaagacacttggcgaagagactgccccaga 2112
|||||
Db 2041 ACAAGAGCAAGTGTCGATTTTAAAGAAAAAGACACTGGGCAAGGAAAGTCCCAAGA 2100
QY 2113 agggaacaaagagaccgaagtcctagctctagaagaagataaagattaggggagaggg 2172
|||||
Db 2101 AGGGAACAAGAGACCGAA-GTCTTAGCTCTAGAAAGAGATAAATTTAGGGGAGAC-GG 2158
QY 2173 gttcggagccctccccggcccaaggttaactltgaagtggaaggggcaacacgtgaat 2232
|||||
Db 2159 GTTCGGAACCCCTCCCGAGCCCAAGGTAACCTTTGAAGGTGAGGGGCAACCAAGTTGAGT 2218
QY 2233 tctctgttga taccgagagcgagcaltcag tgcgtctacaacacaltlaagaaactaaag 2292
|||||
Db 2219 TCTGTGTTATACCGAGCGGAGCATTTAGTGTCTCAACCAATTTAGGAAAAACTAAAG 2278
QY 2293 aaaaaaaatccctgggtgtagtgggtgcacagggcaaggaagfatcatagagctacccga 2352
|||||
Db 2279 AAAAAAATCTGTGATGAGGTCACAGAGGCAAGGAGATATCATATGATACCGGAA 2338
QY 2353 gaacgttgacttggaggtgtagacgg tlaaccactcgtltctgtatcccttgagtgcc 2412
|||||
Db 2339 GAACCGTTGACTTGAGGTGGAGTGGAGCGGTAAACCACTCGTTCTGGTCATCCTTGAGTGCC 2398
QY 2413 caatccctctaggtlaagagactlactgacaaagatgggagctloaaatttctttgaac 2472
|||||
Db 2399 CATTACCCTTCTGAGTATAGACTTACTGACCAAGATGGAGCTCCAAATTTCTTTGAAAC 2458
QY 2473 aaggaaagcagaagtgctgtgataaacaacccatcaagtggttggcccccataag 2532
|||||
Db 2459 AAGGAAGACAGAGAGTGTGTGAATACAAACCATCATCTGTGTGAACCTCCCAATTAG 2518
QY 2533 atgataaata tgaactatcttccccagtaagcctgatatgaataataacgttcgt 2592
|||||
Db 2519 ATGATGAATATCCACTATATTTCCCCAGTAAAGCTGATCAAGATATACAGTCTGTGT 2578
QY 2593 tggagcagtlcccccagccttggcagaaacccgcaagatgggtlttggaaagcaagttc 2652
|||||
Db 2579 TGGAGCAGTTTCCCAAGCCTGGGCAAGAACCGACAGAGATGGGTTGGGAAAGGATTC 2638
QY 2653 ccccaaggttaltcaacttgaagccaggtgtacacacagatlcagtcagacagtaacct 2712
|||||
Db 2639 CCCACAGGTATTCACCTGAAGCCAGTGTCTACACAGTATCAGTCACAGTACCCCT 2698
QY 2713 tggatgaagagctcgcgaagaaatttggcgcagatgtlcaaaagatlaalccaaagggca 2772
|||||
Db 2699 TGAATAGAGAGCTCGAGAAAGAAATTGGCCGATGTCAAAAGATTAATCCAAAGGGCA 2758
QY 2773 tccatgttccgtlccaa tcccccttggaaatcctccctgtctaccgggtlaagaaagccttgg 2832
|||||
Db 2759 TCTAGTTCCGTCCAAATCCCTTGGAAATCTCCCTGTACCGGTTAGGAAACCTTGGA 2818
QY 2833 ccaatgtatcagcagcagtaacagagcttgaagaggtlcaaaagaggttgcagagatac 2892
|||||
Db 2819 CCAATGATTATCGACCAAGTACAGACTTGAGAGAGGTCAATAAAGGGTCCAGAGCATAC 2878
QY 2893 acccaaggtccgaacccctataacctlttgagcgctcccgcttgaacggaaatgt 2952
|||||
Db 2879 ACCCAAGCTCCCAACCTTATTAACCTTTGACCGCCCTCCGCTGAAGGAACTGGT 2938
QY 2953 acaagatltgacttlaaaga tgccttctctgcctgagatlaacacccactagcaac 3012
|||||
Db 2939 ACACAGTATTTGACTTAAAGATGCCCTTCTCTCCGAGATTACACCCACTAGCCAAC 2998
QY 3013 caacttctgcctcgaatgtgagagatccaagtacggaagaagacccggcagctcaactgga 3072
|||||
Db 2999 CACTTTTGGCTTCAATGAGAGATCCAGGTAGCGGAGAAACCGGCACTCACTCGGA 3058
QY 3073 cccgacgtcccaaggttcaagaaactcccgacactcttgaagagccctacacaggg 3132
|||||
Db 3059 CCGGACTGCCCAAGGGTTCAGAAACTCCCGACCATCTTTGAGAGAAACCTTACAGGG 3118
QY 3133 acctggcgaacttcaagatccaacacccctcaggtgacccctccactcaatgaatgaac 3192
|||||
Db 3119 ACCTGGCCACCTTCAGATCCACACACCTCAAGGTGACCTCTCCAGTACGTGATGACC 3178
QY 3193 tgccttgcggagagccacaaacagagactgtcttagaaaggttaagaaagcatalactgttg 3252
|||||
Db 3179 TGCTTCTGGCGGAGGACCAACAAAGAGACTGTGTAAGAGTACGAAGGCACTACTGCTGG 3238
QY 3253 aattgtctacttagctcagagcctctgctcaagaaagcccaatlttgcagagagag 3312
|||||
Db 3239 AATTGTCTACCTAGGCTACAGAGCTCTGCTAAGAAAGCCCAAGTTTGAGGAGAGAGG 3298
QY 3313 taacatacttgggtlcaaglttgcgggagcgagcagatgtctgacggaagcaggaaga 3372
|||||
Db 3299 TAACTACTTTGGGCTACACTTTGGGGGCGGCGACGAGTGGCTGACGAGGACGAGAA 3358
QY 3373 aaactgagtccagataacggcccaacacagacgaacaaacaaatgtgagagatlttgggga 3432
|||||
Db 3359 AAACGTGATTCAGATACCGGCCCAACACACAGCCAAACAAATGAGAGAGATTTGGGGA 3418
QY 3433 cagctgagatlttgcagacttggatcccggggtlttgcgaacttagcagcccaactctac 3492
|||||
Db 3419 CAGCTGATTTTTCAGACACTGTGATCCCGGGGTTTGCAGCTTAAACACGCCACTCTAC 3478
QY 3493 cgttaaccaaagaaagggagatcttcttgggtctcttgagcaacgaagagcatltggt 3552

Db	3479	CGTAAACCAAAAAAGGAGATTCCTCGGCTCCAGACACAGAGGACATTTATG	3538
QY	3553	ctatcaaaaggccctgctgagcgacactgctgcccctccctgaagtaataaccct	3612
Db	3539	CTATCAAAAAGGCGCTGCGAGCGCACTGCTGGCCCTCCCTGACGTAATCAACCT	3598
QY	3613	ttacccttattgtagtaagcgttaaggagtagcccgaggaggtttaaccaaaccttag	3672
Db	3599	TTTACCTTTTATGTGATGATGAGCTTAAGGAGTAGCCGAGGAGTTTAAACCAACCTAG	3658
QY	3673	gaaccaaggagagacactgtgccaactgctcaaaaggagcttgatccctgtagcaagtggt	3732
Db	3659	GACCATGAGAGAGACCTTGTGCTACCTGTCAAGAGCTTGTATCCGTGAGCCAGTGTT	3718
QY	3733	ggccgctatgctctgaaagcctacgacgctgagcactgctgcaactgctcaagagcgtgacaat	3792
Db	3719	GGCCGTAATGCTGAGGCTATCCGACCTGTGGCCTATGTCGTCAGAGAGCGCTGACAAAT	3778
QY	3793	tgactttggagacagataaactgtagtaagccccccctgcatgtaggaacatgctgagc	3852
Db	3779	TGACTTTGGAGACATATTAATCTATATAGCCCCCATGTCATTTGAGAACATGCTGCGC	3838
QY	3853	agccccagacccgaaggagacacaccccgcaaggacccactaacaagcgtgcttcta	3912
Db	3839	AGCCCCAGACCCGATGATGATGACCAAGCCCGCATGACCCATATCAAACTGCTGCTGA	3898
QY	3913	cagagagggtcacttgcctccacacgcccgtctcaaccctgccaactctctgcttgaag	3972
Db	3899	CAGAGAGGTCACCTTTCGCTCCACAGCCGCTCTCAACCTGCTCTTCTGCTGAG	3958
QY	3973	agactgtagaacagtagactgactgactgcaactaactatgtagtaagagactgagc	4032
Db	3959	AGACTGATGAAACAGTAGCATGATTTGCCATCACTATTTGATGAGGAGACTGGGGTCC	4018
QY	4033	gcaagagaccttaagacataacgcgtgactgtagaagtgctaaactgctgactgacgaa	4092
Db	4019	GCAAGAGACCTTACAGACATACCGCTGACTGAGAGAGTGTAACTGTTCACTGACGGA	4078
QY	4093	gcacgactgtgtggaaggaaggaagtagctggggcgccgctggtggaagggacccga	4152
Db	4079	GCAAGTATGTGGTAGAGTAAAGAGATGAGTGGGGGCGGAGTGGTGGACGGGCGCA	4138
QY	4153	cgactgggacgacgctgcccgaaggaactcaagcgaagcgtgagcttagctgccc	4212
Db	4139	CGATCTGGGCGCAGCAGCCGCGGAGGAACTTCAAGCCCAAAAGGCTGACTGACCC	4198
QY	4213	taacgaagccttgagcgtgcccgaaggaataccataaacttalaagacagcaggt	4272
Db	4199	TAACGCAAGCTTTGGGCTGGCGCGAAGGAAATCCATTAACATTTATACGACAGCGT	4258
QY	4273	atgctctgagctgacacagtlacaggggcatctataaacaagggttgcttaact	4332
Db	4259	ATGCTTTGGCAGCTGACACAGTACACGGGGCATCTTTAAACAAGGGGGTTCCTTAACCT	4318
QY	4333	cagcggagggagaataaagaacaaggaatctaaagcctatagaagccttcaat	4392
Db	4319	CAGCAGGAGGGAATTAAGAAAGAGGAATTTAAACCTATTAGAAAGCTTTACATT	4378
QY	4393	tgccaagaagctagctattataactgcttctggaactcaagaagccaaagatctaat	4452
Db	4379	TGCCAAAAAGGCTAGCTATTATACACTGTCTTGACATCGAAGGCAAGAGCTCTATAT	4438
QY	4453	ctagaagggaacagatgctgacccgggttgccaagcagagcagccagcgtgttaacctc	4512
Db	4439	CTAAGAGGAACCAAGATGGCTGACCGGTTCCAAAGCAGGAGCGCCAGGCTGTTAACCTTC	4498
QY	4513	tgctaatatagaacccccaagcccccgaaacccagacgacgtaacacctagaagct	4572
Db	4499	TGCTTATATATGAACCCCAAGGCCCAAGAACCCAGACAGATACACCTAGAAAGCT	4558
QY	4573	ggcaagagataaagaatagacagttctctgagactcgaaggagactgctataact	4632

Db	4559	GGCAGACATATAAAGATAGACATTTCTGTGAGACTCCGAGGGGACCTGTATACCT	4618
QY	4633	catatgggaaggaaatccctgcccacaaaggaggttagaataatgctcaacagataac	4692
Db	4619	CATATGGGAAGAAATCTGCCCCACAAAGAGGTTTAGATATGTCTCAACAGATCATC	4678
QY	4693	gtctaacccacctagaagaaactaacaactgacagctgctgctcagaacatcccttataatg	4752
Db	4679	GTTTACCCACCTAGGAACCTTAACACCTGCAGCTGAGTGGTCAGAACATCCCTTATCATG	4738
QY	4753	ttcttaggctacccggagtgctgactcgggtggttccaacttggtgcccctgacgctg	4812
Db	4729	TTCTGAGGCTACAGAGACTGCTTACTGCTGGGTCAAACTTTGTGCTGCTGCACTGG	4798
QY	4813	ttaatgctaatccctccagaataactccagaagaagactcaagggaagccacccaggg	4872
Db	4799	TTTATGCTAATCCCTTCAGAAATACCTTCAGAAAGAGACTAAGGGAGACCCAGGCG	4858
QY	4873	ctcaactgggaagtgagcttaactgagtgtaagccggcctaatacggaaacaaatactat	4932
Db	4859	CTCAGTGGGAATGAGACTTCACTGAGGTAAAGCCGCTAAATACGAAACAAATTTCTAT	4918
QY	4933	tggttttgtagaacacttccaggaaggtaggaagctataactctaaagaagaagact	4992
Db	4919	TGTTTGTGTAAGACCTTTTCAGATGGGTAGAGCTTATCTCTAAGAAAGAGACTT	4978
QY	4993	caaccggtggtgctgaagaaatactgtaggaagaaatttccaagatttgaaatacctaag	5052
Db	4979	CAACCGTGGGTGCTAAGAAATACTGAGAAATTTTTCAGATTTTGAATTTGTAAG	5038
QY	5053	taaataggtaagacaaatgctccagcttctgttcccaagtaagtaaggactgagcaaga	5112
Db	5039	TAAATAGGTGACGACATTTGCTCAAGCTTTGTCGCCAGGTAAAGTAAAGGACTGGCAAGA	5098
QY	5113	tattgggagattgattggaactgcatgtgtcaacagaccacaaagcctcagagacgttag	5172
Db	5099	TATTGGGAGTTTATTGGAACACTGTTGCTATGATACAGACCCCAAGCTCAGAGATAG	5158
QY	5173	agaaggatgaataagacccttaagagacccttaacaaattgacacagagactgcaata	5232
Db	5159	AAGAGATGATTAAGAACCATTAAGAGACCTTACCAATTTACACAGAGACTGGCATTA	5218
QY	5233	atgattgtagtgcttccctgccccttgctgtctttaaaggtagaagaaacccctggaagct	5292
Db	5219	ATGATTTGATGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	5278
QY	5293	ttgggctgaacccctcctaagattgctctacaggggagaccccccgttgtagaagaattgct	5352
Db	5279	TTGGGCTGAACCCCTATTAATTTGCTTAAGGGGAGCCCCCGTTGGAGAAATTTGCC	5338
QY	5353	ttgcaacagtgctgtagtgctgcttccagagccttgcttcttaggctcaagcgctcg	5412
Db	5339	TTGCAATATAGTGCATGATGCTGCTGCTTCCAGGCTTTGTTCTTAAGCTTCAAGGCGCTCG	5398
QY	5413	agtggtgtagagcagcagcgttgaagcagctccgggagcctcactcaggaagagacttgc	5472
Db	5399	AGTGGTGAGGACAGCAGGAGCTGGAAGCAGCTCCGGGAGGCTCTACTAGAGAGACTTGC	5458
QY	5473	aagttccacatcgcttccaaagttggagattcagctatglttagaagcgaacgttgaagaa	5532
Db	5459	AAGTTCCACATCGCTTCCAAAGTTGAGATTCATGATTTAGACCCACACCAAGGCTGTGA	5518
QY	5533	acctgagactcgggtggaaggaactatctgtaactttggaacacacaaagagtgta	5592
Db	5519	ACCTGAGACTCGATGGAAGGAGCTTATCTGTAATTTTGAACCAACCAAGGCTGTGA	5578
QY	5593	aagtgaaggaatcccaacttgatccatgcatcccaogtlaagccggcgcaactcccg	5652
Db	5579	AAGTGAAGGAATCCCTACCTGATCATCCACAGCTTAAGCGGCGCACACTCCCG	5638
QY	5653	actcgggttgaagacggaagactggaatcccttaagcttgcctcatcgctcg	5712
Db	5639	ACTCGGGGTGAGAGCGGAAGAAAGACTGAGATTCCTTTAAGCTTGCCTCATGCGCTGG	5698


```

/protein_id="CAB65339.1"
/db_xref="GI:6688946"

```

QY 61 aagacgcctctcgtgagtgattcggggtctgcgcctcttcgcgagccgcgaggggac 120
|||||
Db 627 AAGACCCCTTCTCGTGAGTGATTTGGGGTGTCCGCTCTTCCACCCCGCCACCCCGCCGCG

OY	121	tgtctcttacttactgaccttcaatttgtagtgcgttgccgggaatactcgcgaccacctta	180
Db	687	TGTTCTTTTACTGGCTTTCATTTTGGTGCCTTGGCCCGGAAATTCGCAACCACTCCCTTA	746
OY	181	caaccggaaacgcacttggaaagtaaggaatcccccttggaaacgtgtgtgtgcggcc	240
Db	747	CACCCGGAACCCACTTGGAGGTAAAGGATCCCTTTGGACACTGTGTGTGTCTCGGCC	806
OY	241	ggctccctcttcttgtagtgccttcttcgtatgcgcgccttccgatttgcagctgcct	300
Db	807	GGCCTCTCTTCTTGAGTCTGTGTTTTCGGTGATGCGCGCTTTCGGTTTGCAGCTGTCTCT	866
OY	301	ctcagacccgtaaggaacttggaggactctgtatcaagaacgtgtcaggaggaatacaagcgctg	360
Db	867	CTCAGACCCGTAAAGACTGGAGACCTGTATCAGAGACGTGTCAAGGAGATCACAGCGTG	926
OY	361	ccaccctggggagcgcgcccgaggcttggggagagcagagagcgcctgtgtctctact	420
Db	927	CCACCCCTGGGGAGGCCCGGGAGGTGGAGAGACCGAGACGCTGTGTGTCTCTACT	986
OY	421	gtcgtcagaaggagacgagctcgttctgtttgaagcgaaagcttccccctccgcgcgctcg	480
Db	987	GTCCGTGAGAGAGACCGAGTCTGTGTTTGAACCGAAACTTCCCCCTCCGCGCCTCGCG	1046
OY	481	actcttctgcctgtcttltggaagacggagagcgtgcgtgtgtctgtatcttgttgttt	540
Db	1047	ACTCTTTTGGCTGTGTTGGAAACCGCGAGCGGTCTGTGTGTCTGTATCTTTGGTTT	1106
OY	541	ctgtctcgtgtgtcttctgtcttltgtgtgcctctgtctacagttttaaataatggaacaag	600
Db	1107	CTGTTCCTGTGTGTGTGTCTTGTGTGGTCTCTTGTCTTACAGTTTAATATGGAGACAGAG	1166
OY	601	tgaatacccccttagtttactctcgcacatcttgacactggaactgaatagatcaagagctcata	660
Db	1167	TGACGACCCCGCTTAAGTTAACTCTGACACTGAGATTGAGATTACAGAGGCTCTATA	1226
OY	661	atttgcagttcaggttaagaagggaccttggcagacttcttgcctcttgaatgggcaaa	720
Db	1227	ATTTGTCTAGTTTCAGTTTAAGAAGGGACCTTTGGACCTTTGTGCTCTGTAAATGGCCAA	1286
OY	721	catctgattgttggcttggtgcacatcagagggagacctaattctbaatactctgtcgttla	780
Db	1287	CATTGAGATTGGATGGCCATCAGAGGGACCTTTAATTCGAAATTATCTGGCTGTTA	1346
OY	781	aggcaataattttcagacttggaccgycgtctcatccttgaatcaaggagccatactctta	840
Db	1347	AGGCAATCATTTTTCAGACTGTGACCCAGCTCTCATCTGATCAGAGACCCCTATATCTCTTA	1406
OY	841	cgttgcagaatttggcagaagaatccctccgcatacgtgttaaccatcgttataataaacaa	900
Db	1407	CGTGGCAATCTTGGCACAAGATCTCTCGGCATAGGGGTTAAACCATGTGCTTAATTAACCAA	1466
OY	901	gaaagcaggtccccaatctctgtctcttggagagaaaaaacaacactccgcgcgaaag	960
Db	1467	GAAAGCCAAGTCCCGAATCTCTGGCTTTTGGAGAGAAAAACAACATCTGGCGAAAAAG	1526
OY	961	tcaagcccttctcctcgtatctaacccgagatcagagagccgcgcgaacttggcgcgaacc	1020
Db	1527	TCAAGCCCTC-TCCTCTGATCTACCCGAGATCAGAGAGCGCCGCACTTGAGCCGGAACCC	1585
OY	1021	caactgttctccccaacccccctatccagcaacaggtgtcgtgtgaaggga--cctctgtccct	1079
Db	1586	CAACCTGTTTCCCCACCCCTTATTCACAGCACAGGCTCTGTGAGGGAGACCTCTCTCCCT	1645
OY	1080	ccctggagctccggttgttggaaagacactctgcgcggagctcgtggagccggagagcgcaac	1139
Db	1646	CTGTGAGCTCCGGTGTGTGAGGGACCTTCTCTCCGGACTGTGGAGCCGGAGAGCGCCACMC	1705
OY	1140	ccggagcggaacaacagatctcgatatactacgcttcgcacatactgacctccaatgcga	1199
Db	1706	CCGAGACGGGACAGACAGATTCGCGATTATTAACGCTGGCGACCTATGCGCTCCCATATCGCG	1765
OY	1200	ggggggccaatttgcagccctccatgattgcaccttctctctgcgagatctctataattgg	1259

Db	1766	GGGGGCAATTGACAGCCCTCCAGATTGGCCCTTTCTCTGCAATCTCTAATAATGG	1825
Qy	1260	aaactaacatcccccttctcggagatccccaacgcctcacggygttggatctc	1319
Db	1826	AAACCTAACATCCCTCTTCGAGGATGCCAACGCCCTCAGGGGTGGGAGATCC	1885
Qy	1320	cttatgtctctaccagcctacttggatgatgtcacacagctgtctcagacatcttc	1379
Db	1886	CTTATGTCTCTCACAGGCTCTTGAGATATTGTCAACAGCTGCGAGACCTCTTC	1945
Qy	1380	acaacgaagagcagagagaaattctgttaaggtctgaaaaatgtctcttgagccgac	1439
Db	1946	ACAACCGAGAGGAGAGAAATTCGTTAAGGGCTGAAAAAATGTCTCTGGGGCGAC	2005
Qy	1440	ggagagaccacgcagttgcaaatagattgacatgagattcccttactgcgccggt	1499
Db	2006	GGGGACCCACGACGTTGCAAAATGAGANTGACATGGGATTTCCCTTGACTCGCCCGGT	2065
Qy	1500	tggaactacacacagcgtctgaagtgagagagcttgaaaaactctatgcagcgtctcgtg	1559
Db	2066	TGGGACTACAACAGCGCTGAAAGGTAGGGAGACTTGAAAACTATGCCCGAGGTCTGGTG	2125
Qy	1560	gcgggtctcggggcgctctaagaaacggcccactaaattggcttaaggtcaaaagagtgatg	1619
Db	2126	GGGGTCTCCGGGGCGGCTCAAGACGGCCCACTAATTTGGCTAAGTAAGAGGTGATG	2185
Qy	1620	caggagaccgaacgaacctccctcgttattctcttgagagctcatalygagccttcaggg	1679
Db	2186	CAGGGACCGAAGCAACTCCCTCGGTATTTCTTGAGAGGCTCATGGAAGCCTTACAGCGG	2245
Qy	1680	ttcaccccttltga tctactaccagagcccgcaaaagcctcagtgagccctggtcttcat	1739
Db	2246	TTCCACCCCTTTGATCCTACTCTCGAGGCGCCAGAAAGCCTCAGTGCCCTGCGCTTATAT	2305
Qy	1740	gggcagctcgctctcgtatcacgagaagaactcagagactggaaggtttacagaaagct	1799
Db	2306	GGGCGATGCGCTCGGATATCAGAAAGAACTTCAGAGCTGGAAGGTTTACAGAGGCT	2365
Qy	1800	ggattacgttcatcagtgagagagcagagaaagtgtattacagaagggagacagaag	1859
Db	2366	GAGTTACGTGATCTAAGTAGAGAGCCAGAAAGGTATTACAGAAAGGAGACAGAAAGAG	2425
Qy	1860	gagaaagaaacagagaagaagaagagagaagaaagggagagaaagcgtgatcagcg	1919
Db	2426	GAGAAAGAAAGACAAAAAGAAAGACAGAGAAAGGAGGAAAGCGTGTATGACGG	2485
Qy	1920	caagagaagaatttgctaaagatcttggccgagtggttggaagggaagaagcagcagggag	1979
Db	2486	CAGAGAAATAATTTACTAATAATTTTGCCGCGAGTGTGAAAGGAGAGACGACAGGGAG	2545
Qy	1980	agagagaagatcttagaanaattaaagtcagagccctacagatcagggaaacctggacaat	2039
Db	2546	AGGAGAGAGATTTTAGSAAATTAAGTCAAGCCGTACAGACGTACAGGAAACCTGGCGAAT	2605
Qy	2040	aggaaccacactcgacaagagaccagtgtgcatctgttaaagaanaagagacatgggcaag	2099
Db	2606	AGGACCCCACTCGACAAGGACACAGTGTGCGATTGTAAACAAAAAGACACTGGGCAAGG	2665
Qy	2100	aactgcccagaagaggaaacaagaagccgaagttctctagctctaaagaagaataaagat	2159
Db	2666	AACTGCCCAAGAGAGAAACAAGAGACCGAAGGCTCTTAACTTTAGAAAGAAATTAAGAT	2725
Qy	2160	taaggagagaggggttcgagacccccctcccgagcccgaggtlaacttgaagtgagggg	2219
Db	2726	TAGGGAGAGAGGGGTTGAGACCCCTCCCGAGCCAGGGGTAACTTTGAAGGTGAGAGGGG	2785
Qy	2220	caaccagttgagttctctgttgtalacacggagcggagatcagtgctgtctaaaccatta	2279
Db	2786	CAACGATTTGATTCCTGTGTGATACCGGAGCCGAGCATTCAGTGTGCTTACAAACATTA	2845
Qy	2280	ggaanaactaaagaaaaaaatccctgggttgatgggtgcacagggcaacggcagatcca	2339

Db 2846 GGAAACCTTAAAGAAAAAATCTGGGTGATGGGTGCGACAGGGGACAGCGCATATCCA 2905
Qy 2340 tggactaccggaagaccgttgaacttggagtggaacgggtaaccactggtttcgtc 2399
Db 2906 TGGACTACCCGAGAAACCGTTGACTTGGCACTGGAGCGGGTAACCCCTGTTCTGTGTC 2965
Qy 2400 atcccttgatgcccgaatccccttctagtagagacttactgaacaaatggagctcaa 2459
Db 2966 ATCCCTGAGTCCCAAGTAACCCCTTCTAGTAGACACTTACTGACCAAGATGGAGGTCAA 3025
Qy 2460 attctcttgaacaagaagaccgaagtgctcttgaalaaacacccatctgtgttg 2519
Db 3026 ATTCTTTTGAACAAGAGACCGAAGTGTCTGTGATATACAAACCCATCCTGTTGTTG 3085
Qy 2520 accctcaatagatgtaalaaalaaalaaalaaalaaalaaalaaalaaalaaalaa 2579
Db 3086 ACCCTCCAAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3145
Qy 2580 atacagctcgtgttgagagaggttcccccaagccttggcagaaacgcagagatgggttg 2639
Db 3146 ATACAGTCTGTGTGGAGAGAGTTCCTCCCAAGCTGGGCAAAACCCAGGGATGGGTG 3205
Qy 2640 gcaaaagaaagttcccccaaggtatlatcaactgaagcagtgctacacacatlatcagtc 2699
Db 3206 GCAAAAGCAAGTTCCTCCCAAGGTTATTCACACTGAGGCCAGTGTCTACCATATCAGTC 3265
Qy 2700 agaaagtaaccccttggatagaagagctcgagaagaaatttggccgcatgttcaagaatla 2759
Db 3266 AGACAGTACCCCTTGAATGATGAGAGGCTCGAAGAGAAATTTGGCCGATGTTCAAGATTA 3325
Qy 2760 atcaaaagagacatcctagcttctgtccaatcccttgaatccttccctgttaccgtt 2819
Db 3326 ATCCAAAGGAGATTCAGTTCCTGTGTCATCCCTTGAATATCCTCCTGTACCGGTT 3385
Qy 2820 aggaagccttggagacaaatgattatcgaaacagagacttgaagagagtcataaag 2879
Db 3386 AGGAAGCTTGGAGCAATATATTCAGACAGTACAGAGACTTGAGAGAGTCAATTAAGG 3445
Qy 2880 gtgcagagacatacaacaaacggtcccgaaacccctttaaacttgaagccctccgcct 2939
Db 3446 GTGCAGAGACATACACCAACCGGTCCGAAACCTTTATTAACCTTTTAAAGGCCCTCCCGCT 3505
Qy 2940 gaacggaactgtgtacacagatatttgaactaaagaatgctcttctgtcctgagattacac 2999
Db 3506 GAACGGAGACTGGTACACATATTTGACTTAATAAGATGCTTCTGTGCTGAGATTACAC 3555
Qy 3000 cccactaagcacaacacacttcttgccttgaatggagaagatccaagtaagggaaacccgg 3059
Db 3566 CCCACTAGCCAAACCGCTTTTGCCTTCGAAATGAGAGATCAAGGTACCGGAAGAACCGGG 3625
Qy 3060 caagctaaccttgaacccagactgcccgaaggttcaagaactcccgaaacatcttgaagaa 3119
Db 3626 CAGCTACCTGAGACCGGACTGCCCCAAGGTTCAAGAACTCCCGACCATTTTGAAGAA 3685
Qy 3120 ggccttaacagaagacttggccaacttcaagatccaacacccctcaagtgaacccctccacg 3179
Db 3686 GGCCTTACACAGGAGCTGGCCAAATTCAGAGTCCAAACCCCTCAGGTGACCTCTCCACG 3745
Qy 3180 taacttgaactgtcttctgacggaagcccaacaaagagacttcttaagaagtaagaa 3239
Db 3746 TACGTGATGACCTGCTTCTGGCGGAGACCAACAAAGGACTCTTAAGAAAGGTACAAAG 3805
Qy 3240 gaactaactgttgaatgttctgaacctaagctacagagccttctgctaaagaagcccaatt 3299
Db 3806 GCACACTGCTGGAATTTCTGACCTAGGCTACAGAGCCTCTGTAAGAAAGGCCCAAGAT 3865
Qy 3300 tgcagaagagaaggttaacacttgggttacaggttgcggggcgagcagaatggctgacg 3359
Db 3866 TGCAGAGAGAGGTAACTTGTGGGTACAGTTTGGGGGCGGAGCATGGCTGACG 3925
Qy 3360 gagagacggaagaaactgtatcagataccggcccaacacacagcaagaagtgaga 3419
Db 3926 GAGGACCGAAGAAAGTGTAGTCCAGATACCGGCCCAACACAGCAAGCAAGTGA 3985
Qy 3420 gaatttttgggagacgctggaatttgcagactgtgataccggggttgcagacttaaga 3479
Db 3986 GAGTTTGGGAGACAGCTGATTTTGCAGACTGTGATGCCGGGTTTGGACCTTAGCA 4045
Qy 3480 gcccaactctacccgcttaaccaagaagaaggagatcttcccttgggtccttgaacacag 3539
Db 4046 GCCCAGCTCTACCCGCTTACCAAAAGAAAGGGAATTCCTGGGCTCCTGAGCACAG 4105
Qy 3540 aagcatttgaatgctatcaaaaagccctgctgagcagccttcttggccctcccttgaac 3599
Db 4106 AAGCATTTGATGCTATCAAAAAGGCCCTGAGGACACTCTCTGACCTCCCTGAC 4165
Qy 3600 gtaactaaaccccttacccttcttgaatgagctgaagagtagcccgagagtttla 3659
Db 4166 GTAACCTAAACCCCTTACCTTTATGTGATGAGACCTGTAAGGAGATGCCGAGAGATTTA 4225
Qy 3660 acccaaacccctlagagacatgagagagaccccttgccttactcttcaagaagcttgaact 3719
Db 4226 ACCCAAAACCTAGAGACCATGAGAGACCTGTTGCTACCTGTCAAAAGATGATCT 4285
Qy 3720 gtagcagtggttggccgtagtctgaagctatcgcaagcttggccatctgag tcaag 3779
Db 4286 GTAGCCAGTGTGGCCGCTATGCTGAAGGCTATGAGCTGAGCTGTGGCATATGTTGACAG 4345
Qy 3780 gaacgtgacaacttgaacttgggacagaaatataactgtataagcccccaatgag 3839
Db 4346 GACGCTACAAATTTGACTTTGGAGCAAGATATTAATGATGATGATGATGATGATGATGAT 4405
Qy 3840 aacatgcttgcagagcccccagacagatgatagaacaaagcccgatgacacacataa 3899
Db 4406 AACATGTTTGGCAGCCCCAGACGATGATGATGATGATGATGATGATGATGATGATGAT 4465
Qy 3900 agccttcttcaagaagaggttcaacttgccttcaacagccgcttcaacccctgcaact 3959
Db 4466 AGCCTGCTTCTCAAGAGAGGCTCACTTTCCTCACCAGCCGCTTCACCTGACACT 4525
Qy 3960 ctcttgcctgaagaagactatgaacagtgatactgttgcatacaacttgaatgag 4019
Db 4526 CTCTGCTGAGAGAGACTATGATGACAGTATGATGATGATGATGATGATGATGATGATGAT 4585
Qy 4020 ggaactggggttccggaagacctaagacataaccgctgaacttgaagaagtgttaacctg 4079
Db 4586 GAGACTGGGGTCCGAAAGACCTTACAGACATACCGCTGAGTGAAGTCTAACTGAG 4645
Qy 4080 ttaactgaacggaagcaactgttggtagaaggttaagagatggttggggcgagctgtg 4139
Db 4646 TTCACGTGAGAGAGAGAGCTATGTGTGTAAGGTAAAGATGCTGGGGCGCGGTGTG 4705
Qy 4140 gacgggaacccgacagatctgggccaagcagccttgcggaagaaattcagcgcaaaagct 4199
Db 4706 GACGGAGCCCAACGATGTGGGCGAGCCTGCGGGAAGAACTTCAAGGGCAAAAGCT 4765
Qy 4200 gagctatggcccttgaacaaagcttgggttggcgaagggaaatccataaacttact 4259
Db 4766 GAGCTATGAGCCTTACCAAAAGCTTGGGCTGCGCAAGGAATTCATTAACATTAT 4825
Qy 4260 accgaacagagtagtgccttgcagacttgaacagctacaggggcatctataaacaag 4319
Db 4826 ACGGACACAGGATAGCTTGGCAGTGCAGACGTACACGGGGCATCTTAACAAAG 4885
Qy 4320 ggttggcttactcagcagggagggaaataaagaacaagaagaaattcctaactatla 4379
Db 4886 GGGTGTCTTACTCAGCGAGGGAAGTAAGAAACAAAGAGAAATTTCAACCTATT 4945
Qy 4380 gaagccttaacttggcaaaaagcttagctattataaactgttcttgaatcagaagcc 4439
Db 4946 GAAGCTTACATTTGGCAAAAAGCTTGTATTAATACGTCTTGACATCAAGAAAGCC 5005
Qy 4440 aaagatctcalatcagaaggaacagatgcttgaacgggttcccaagcagggagccag 4499
Db 5006 AAAGATCTATATCTAGAGGAGAACAGATGCTGACCGGGTTCACAGCAGGACGCCAG 5065

OY 1320 ctatgtctctcacagctacttgatgtatgtcaacagctgctcagaactctc 1379
|||||
Db 1847 cttagtctctcaccagcttacttggatgtgtcaacagctctcagacactcttc 1906
OY 1380 acaacgagagcagagagaaatctgttagagctagaanaatgcttcggggccac 1439
Db 1907 ACAACCGAGAGCGAGAGAAATCTGTAGAGGCTAAGAAAATGTTCTCGGGCCAC 1966
OY 1440 ggggaccacagcagtttcaaatgtatgtacatggatctccctgactgcgcggt 1499
|||||
Db 1967 GGGCAACCCACGAGTTCCAAATGATGACATGGATTTCCCTTACTCGCCCGGT 2026
OY 1500 tggagactacaacagcgtgaaagttagggagagcttgaanaatcttcgcagctggtg 1559
Db 2027 TGGGACTACAAACAGCGCTGAAGGTAGAGAGCTTGAANAATCTATCGCCAGGCTCTGTG 2086
OY 1560 ggggtctccgggagcctcagaagcgccactaatctgtgctaaggttaagagatgtg 1619
Db 2087 GCGGGTCTCCGGGGGCTCTCAAGACGGCCACTAATTTGGCTAAGGTAAAGAGGTGATG 2146
OY 1620 cagggaaccgaacgaactccctcggtattcttgaagagctcatgtgaagcctcagcg 1679
|||||
Db 2147 CAGGAGCCGAGCAACCTCCCTCGTATTTCTTGAAGGCTCATGGAAGCCTTCAGGCGG 2206
OY 1680 ttaaccccttga tcttaacctcagaagcccaagaagcctcagtggtgcctgctcat 1739
|||||
Db 2207 TTCAACCCCTTTGATCTACTACTAGAGCCCAAGAACCTCAGTGCCCTTGCCCTTCAT 2266
OY 1740 ggggaagtcggtctggtatcagaagaanaactcagaagctggaaggtttacagagct 1799
Db 2267 GGGGAGTCGCTGTGATATCAGAAAGAACTTACAGAGTGAAGGGTTACAGAGGCT 2226
OY 1800 gagttagctatctagtagagagagcagagaaggttatcaagaagggagagaagaag 1859
Db 2327 GAGTTAGCTATCTAGTGAAGAGCGCAGAGAAAGTGTATTACAGAGGGAGACAGAAAG 2386
OY 1860 gagaaggaacagagaanaagagagagagaagaagaaggaagacgtgtatagagc 1919
Db 2387 GAGAAAGGACAGAGAAAGAAAGAGAGAGAAAGGAGGAGGAGGAGGAGGAGGAGG 2446
OY 1920 caaggaagaagatttgaactaatcttgcccgagtggtgtgaaggaagcagcagagag 1979
Db 2447 CAAGAGAGAAATTTGATTAAGATCTTGCCGCGAGTGTGAAGGAGGAGCAGCAGGAG 2506
OY 1980 agagagagagattttaggaanaattagtgacagccctagacactcagagaaactggcaat 2039
Db 2507 AGAGAGAGAGATTTTAGGAAATTTAGGTAGGCCCTTAGACAGTCAAGGAACTGGGCAAT 2566
OY 2040 aggaacccactcagaagaagcagtggtgtatctgtaagaanaagacactgggcaag 2099
Db 2567 AGGACCCCACTCGACAAAGAGCAGTGTGCTATTGTAAGAAAGAAAGGAGACTGGGCAAG 2626
OY 2100 aactgccccaaagaagaagaagaagaagcagaggtctcctagctctgaagaagaataagat 2159
Db 2627 AACTGCCCCAAAGAGGAGAAACAAAGAGCCGAAGGTCCTAGCTCTGAAGAAATATAAGAT 2686
OY 2160 taggggagacggtgtcgaacccctcccgagcccgaggttaacttgaaggttagaggg 2219
Db 2687 TAGGGGAGACGGGTTTGGAGCCCTCCCGAGCCCAAGGTAACTTTTAAAGGTGAGGGG 2746
OY 2220 caacagattgagttccctggtgtatcagcgaagcgaactcaatgtctgctaacaactta 2279
Db 2747 CAACAGATTGAGTTCTGTTGATACCGAGCGAGCATTCAGTCTCTCAACCACTTA 2806
OY 2280 ggaanaactaaagaanaaataactcgtggtgtatgtgccaagggcaagcagatcca 2339
|||||
Db 2807 GGAANAACATAAAGAAATAAATCTCGGGTGTATGGGTGCAACAGGGCAACGCACTATCCA 2866
OY 2340 tggactaccgaagaacggttgacttgagtggaacggttaacccactgcttctgtgc 2399
Db 2867 TGGACTACCGAAGAACGTTGACTTGGGAGTGGAGCGGTAAACCCACTCTTTCTGTGTC 2926
OY 2400 atccctgagtgcccaagtaaccctctagtagagacttaactgacaaagatlgagaccaa 2459
|||||
Db 2927 ATCCCTGAGTGCCAGTACCCTCTTGATGAGACTTACTGACCAAGATGGAGCTCA 2986
OY 2460 attcctcttgaacagaagaacagaagtgctgtgaataacaaacactcaactgtgtg 2519
Db 2987 ATTTCTTTGAGACAGGAAACACAGAAAGTGTGTGTAATTAACCAACCACTCACTGTGTG 3046
OY 2520 accctccatctagatgtatgtactatcttccccaaagtaaaagcctgactcaagat 2579
Db 3047 ACCCTCAATTAATGATGATGATATCGACTATATTCTCCCAAGTAAAGCTGATCAAGAT 3106
OY 2580 atcagtcctggttgagcaggtttccccaagccttgggcaagaacgcaaggttgggttg 2639
Db 3107 ATACAGTCTGTTGGAGCAGTTTCCCAAGCCTGGGCGAAGAACCCAGGATGGGTTTG 3166
OY 2640 gcaaaacagttcccccaaggttatctcaactgaagcagtgctacacagatcaagtc 2699
Db 3167 GCAAAACAGTTCCCCACAGGTTATTAACCTGAAGGCCAGTCTACACAGTATCAGTC 3226
OY 2700 agacagtaacccctgtgtagagagcttgagaagaagatttggccgcatgtltaaaagta 2759
Db 3227 AGACATTAACCTTTGATGATGAGAGGCTCGAGAAAGAAATTTGGCCCATGTTCAAGATTA 3286
OY 2760 atccaagggcatcctagttccctgtccaatcccttggataatcctcctgtacaggt 2819
Db 3287 ATCCAAAGAGGATCTCTAGTTCTGTCGCAATCCCTTGAATATCCCTGCTACGGTT 3346
OY 2820 aggaagccttgggaaccaatgatataatgacacaggtacagagacttgaagaagtcataaag 2879
Db 3347 AGGAAGCTTGGAGCAATGATATTATGACCAAGTACAGGACTTGAAGAGGTCAATTAAG 3406
OY 2880 gtcaggaataatacccaacggtccccaaccttaaacctcttgaagccctccgct 2939
Db 3407 GTGCAAGCATATACCCAAAGGTCCCGAACCTTTAATCTTTGAGGCGCCCTCCGCT 3466
OY 2940 gaaaggaactgttacaagatlttgaaatgaatgctcttcttctgtcgtgataaac 2999
Db 3467 GAAGGAATGTGTACACAGTATTGACTTAAAGATGCTTCTGCTGAGATTTACAC 3526
OY 3000 cccactagaaccaacacttcttgccttgaatggagagatccaggttagggagaacggg 3059
Db 3527 CCGACTAGCCAAACCTTTTGTGCTTGAATGAGAGATTCAGGTAGCGGAAAGAACCGGG 3586
OY 3060 cagctaaccttgagcccgactgcccagaaggttcaagaactcccgacacactttagagaa 3119
Db 3587 CAGCTCACCTGAGCCGACCTGCCCAAGGTTTCAAGAACTCCCGACATCTTTGAGAGAA 3646
OY 3120 gacctacacagggagcgtggccaactcaagatccaacacccctcagtgacctctcag 3179
Db 3647 GCCCTACACAGGAGCTGGGCCAATCTCAAGATCCAAACCTCAGTGAGACCTCTCCAG 3706
OY 3180 taagtgaatgacctgtcttgagggagcacaacaaagaagacttgaagaagtagaag 3239
Db 3707 TAGTGATATGACTGCTTGTGCGGAGCCCAACCAAGAGACTGCTTGAAGGTATCCAG 3766
OY 3240 gcaactactgtgaattgtcgaacttagcagagcctctgctaagaagggccagatt 3299
Db 3767 GCACTACTGTGTGAATGTGTGACCTAGCTTACAGAGCCTCTGCTAAGAAAGGCCAGATT 3826
OY 3300 tgcaagagaaggttaaaataacttgggtgtacagtttgcggggcgagcagatgtgtagc 3359
Db 3827 TGCAAGGAGAGGTAAATATCTTGGGGTACAGTTTCCGGGCGGAGGAGTGGTGCAGC 3886
OY 3360 gaggcaaggaagaacactgtagtcagatcacggccccaacacagcagaagaagtaga 3419
Db 3887 GAGGACAGGAAGAAACTGTATGATCCAGATACCGGCCCAACCAAGCAAGAGTGA 3946
OY 3420 gagtttttgggacagctggaatttgcagactgtgataccgggttttgcagccttagca 3479
Db 3947 GAGTTTTTGGGGACACTGTGATTTGCACAGCTGTGATCCCGGGTTTGCAGACTTAGCA 4006
OY 3480 gcccacttaccgggttaaccaaagaagaaggggaattctcctggctctgaagcag 3539
|||||

|||||
Db 3802 TACACTACTGTTGAGTACAGTTTGGGGGCGGACGCTGACGAGGACGAGGAGA 3861
OY 3373 aactctagtcacagataccgagcccaacacagcaagaatgagagatlttgaggga 3432
Db 3862 AACTGTGTGTCAGATACCGGGCCCAACACAGCAAGCAAGTGTGAGAGTGTGGGGA 3921
OY 3433 cagctggatlttgacagacttgagatccggggttgagacttgagccactctacc 3492
Db 3922 CAGCTGGATTTTGGACACTGTGATCCGGGGTTGGCACTTACACACCCACTTACC 3981
OY 3493 ccttaacaaaagaaaggggagatctctcgtggtcctcgtgagacacagagagcatltga 3552
Db 3982 CGCTAACCAAAAGAAAAGGGGAATTCTCGGGGCTCTGAGACACAGAGCATTTGATG 4041
OY 3553 ctataaaaagccctgctgtagcgcaactgtctgcccctccgtagagtaactaacct 3612
Db 4042 CTATCAAAAAGCCCTGCTGAGCCGACCTGTGCTGCTCCCTGACCTTAACCTT 4101
OY 3613 ttacccttatgtgtagtagcgtlaaggagtagcccgagaggtttaaccacacactag 3672
Db 4102 TTACCTTTTATGTGATGAGCTTAAGGAGTAGCCCGAGAGTTTAAACCAACCTAG 4161
OY 3673 gaccatlgagagagacccgtgtgcttaaccgtcaagaagcttgatccgttagccagtg 3732
Db 4162 GACCATGAGAGAGACCTGTGCTGACCTGTCAAAAGAGCTGTGATCTGTAGCACTG 4221
OY 3733 gggccgtatgttgaagagctacgcagcgtgtgcaatactgttcaagaagcttgacaat 3792
Db 4222 GGCCCGTATGTCTGAAGGCTATGCGAGCTGTG6CCATAGCTGTCAAGAGGCTGCAAT 4281
OY 3793 tgactltgagacagataataactgtlaatalagcccccatgacatltgagagacatcgttcgc 3852
Db 4282 TGACTTTGGAGAGATATATACGTATATAGCCCCCATGATGATGGAAGAACTGTCGGC 4341
OY 3853 aagccccaagacagatgagatgacaaagcccgatgacacacttaagaagccgtctcca 3912
Db 4342 AGCCCCAAGCCAGTAGATGAGCAACGCCGATGACCCCATATGAAAGCCGTGCTGCA 4401
OY 3913 cagagagggtcaacttctgctccacagccgctcacaaccttgcaacttctgctgag 3972
Db 4402 CAGAGGGGTCACTTCGCTCCACAGCGCGCTCAACCCCTGCACTGTCGCTGGAAG 4461
OY 3973 agactgtagaacagatgactcatgattgcatcaactatgtattgaagagatctgggtcc 4032
Db 4462 AGACTGATGACCAAGTACATGATGATGCACTTATGATTAAGGAGACTGGGGTCC 4521
OY 4033 gcaagagacttacagacataccgctgactgtagaagtgcttaacctggttcaactgacgaa 4092
Db 4522 GCAGAGACTTACAGACATACCGCTGACTGGAAGTGTCTAACCTGCTTCACTGACGGA 4581
OY 4093 gcaagctatgtgtggaaglaaagagatggtctggggcgagctgtgtggaagaccgca 4152
Db 4582 GCAGCTATGTGTGGAAGTAAAGTAAAGTGTGAGACCGCGGTGAGGAGGACCCGCA 4641
OY 4153 cgaatgggccaagcagccgctggagagagacttcagcgcaaaagcttgagcccaagccc 4212
Db 4642 CGATCTGGGCCACAGCTGCGCGAAGAGACTTACAGCGAAAAGGTGAGCTCATGCGCC 4701
OY 4213 tcaagcaagccttgagcgttgagccgaaggaatccaataacatltatagcagcaggt 4272
Db 4702 TCACGAAAGCTTGGCGGTGGCGGAGGAAATCCATTAACATTTTATAGGACAGAGCT 4761
OY 4273 atgcttctgactgacacgttacaagggccatctataaagaagggttgcttacct 4332
Db 4762 ATGCCCTTTGGACTGCACAGTACACGGGCCATCTATTAACAAAGGGGCTGCTTACT 4821
OY 4333 cagcaggaaggaatlaaagaacaagaagaaatlaagccatlaagagcctlaact 4392
Db 4822 CAGCAGGAGGAAATTAAGAAACAAAGAGAAATTTAAGCCTATTAGAGCCTTACAT 4881
OY 4393 tgcacaaaagctagctatatatacactgtctcgtgacatcagaagccaaagatctcat 4452
|||||

Db 4882 TGCCAAAAGGCTAGCTATTATACACTGTCTGACATCAGAAAGCCAAAGTCTATAT 4941
OY 4453 ctagaaggaaccagatgagctgacccggttgccaagcagcagccagcgttataccctc 4512
Db 4942 CTGAGGGGAACAGATGTGCTGACCGGTTTCCAGACAGGACGCCAGGCTGTAACTTC 5001
OY 4513 tgcctataatagaacgccccaaagcccaagaaaccagacagatcacacctagaagact 4572
Db 5002 TGCTATATATAGAAACCCCAAGGCCAGAACCCAGACAGTACACCTTGAAGACT 5061
OY 4573 ggcagaagataaaaaagatagaccagttctctgaagctccggaggggagactctaacct 4632
Db 5062 GGCAAGAGATTAATAAATAATGACCATGTTCTAGACTCCGGAGGGGACTGCTTAACT 5121
OY 4633 cataatggaagaatctcccccacaagaaggttagaataatgtccaacagatatac 4692
Db 5122 CATATGGGAAGAAATCTCTCCCCACAAAGAGGTTAGAAATGTGCCAAGATACATC 5181
OY 4693 gtctaacccaactaggaactaaacactgcagcagttgtgtcagaacatcccttatactg 4752
Db 5182 GTCTCACCCACCTAGGAATAACACCTGCAGAGTGTGTCAAGAACTCCCTTATCATG 5241
OY 4753 ttctgaagctacaagagtagctgactcgtgtgtgtcaacaactgtgtccctgcaactg 4812
Db 5242 TTCTGAGGCTACCAAGAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5301
OY 4813 ttaatgttaactcctccagataactccagaaagagactaaaggagagccacccagcg 4872
Db 5302 TTATGTCTAATCTTCCAGATATCTCCAGAAAGAAAGACTTAAGGGAGCCACCGCG 5361
OY 4873 ctcaactggagatgagacttaactgaggttaagcccggttaataccgaaataactat 4932
Db 5362 CTCACCTGGGAAGTGACTTCACTGAGGTAAAGCCGGCTTAATACGGAACAAATTCAT 5421
OY 4933 tggcttctgtagaacctlttcaagatcggtagaggttactccactaaagaagaact 4992
Db 5422 TGGTTTTGTAGACACTTTTCAAGATGGGAGAGCCTTATCTTAAGAAAGACTT 5481
OY 4993 caaccgtgtgtgtcctaagaataactctgagagaatlttccaagatlttgaaatccctaa 5052
Db 5482 CAACCTGTGTGCTTAATAAATACTGAGGAAATTTTCCGAGATTTGGATTAATCTTAAG 5541
OY 5053 taataggttcaagacatgctcgaactcgttctgccaagtaagtcaagagactggccaaga 5112
Db 5542 TAACTGGGTGCAAGATGTGCTCACTTTTGTGCCAGGTAAGTCAAGGACTGGCCAGA 5601
OY 5113 taatgggattgtatgtgaactctgcatgtgtcatcagaaccccaagctcagaagtag 5172
Db 5602 TATTGGGATGATTGGAATACTGATGTGTGATCAAGACCCCAAGCTCAGAGACAGTAG 5661
OY 5173 agagatgaatagataccaatlaaagagaccccttacaatltgacacagagactgca 5232
Db 5662 AGAGGATTAATAGAACCTTAATAAGACCCCTTACCAATTTGACACAGAGACTGGCTTA 5721
OY 5233 atgattgtatgtctccctcgtccttctgtcctlttaggtgtgaggaacacccctggagact 5292
Db 5722 ATGATTGATAGCTCTCTCGCTTGTGCTTTTAAAGGTTAGGAACACCCCTGGAGAGT 5781
OY 5293 ttggcttgacccctatgaatgtgtctcaaggaggagaccccccgttggagaaatgtcct 5352
Db 5782 TTGGCTTACCCCTATGAATTTGCTTAAGGGGAGACCCCCCTTGTGTAGAAATTTCTT 5641
OY 5353 ttgcaatagatgactgtgtgttcttccacagccttgttctcctcaagcctcaagcgctg 5412
Db 5842 CTGTACATATAGTGATGTGTGCTTCCAGCTTGTGTCTTGTGTCTTGTGTCTTGTGTG 5901
OY 5413 agtgggtgagcagcagcgtgtgagagcagctccggagagcctactcagg --- aggaact 5469
Db 5902 AGTGGGTAGGCAACGAGACCTGTGAAGCAGCTCCGGAGGACCTACTCTAGGAGAGGAACT 5961
OY 5470 tgcaggtccacatgctctcgaagttagagatlaagctatgtttagagcgccacgctgag 5529
Db 5962 TGCAGTTTCAATCGTCTCAAGTGGAGATGATGCTTATTTAGACGCCACCATGACG 6021
|||||

QY	5530	gaacactcgagactcggltggaagggaacctatctcgtlactctlttgacaacacacgctg	5589
Db	6022	GAACCTCTCGAGACTCGGTGGGAAGGCCCTTATCTGCTACTTTTGAACCAACACAGGCTG	6081
QY	5590	tgaagctcgaaagagatcccccacttgatctcatgcatccacagcttaagccgagccactc	5649
Db	6082	TGAAGTGTGAAGGAATCTCCACCTGGATCCATGTCATCCACGTTAAGCTGGGGCCACCTC	6141
QY	5650	ccgactcgggltgtagagccggaaaagactgagatccccctlaagcttcgcctccatcgc	5709
Db	6142	CCGACTCGGGGTGGAGAGCGGAAAGACTGAGAAATCCCTTAAGCTTGCCCTCATCGCC	6201
QY	5710	tggttccttacttaacaataactccccagccagtagtlaaagccctatagaacgtcgc	5789
Db	6202	TGGTTCCTTACTCTTAACAAATTAACCTCCCGAGGCCAGTAGTAACAGCCTTATAGACAGCTCG	6261
QY	5770	aacccccataagaccttatacccttactcgtcgtatattatgaccccgatagaagggtact	5829
Db	6282	AACCCCATAGACCTTTATCCCTTACCTGGCTATATTATGACCCGTGATACGGGTCTCACT	6321
QY	5830	gtaaatagcactcggagtgltgtccctlaagaggaacctgtgtgcctgaaactgattctgc	5889
Db	6322	GTAATATGCACTCGAGGTGTGTGCTCTAGAGGCACTGGTGCCCTGAAGCTGATTTCTGC	6381
QY	5880	ctccgattgattaaecccgcgtgtlaaagaacacctccacactgctccgtagtataggg	5949
Db	6382	CTCCGATTGATTAAACCCCGCTGTTTAAAGACACCTCCACACCTGTGCTCGTAGTATGCG	6441
QY	5950	tctcattgtcggcccaagcacagagaagaagaatactgtgagggtgttccttgaggaactctc	6009
Db	6442	TTCTATTGCTGCCAGGCACAGAGAAAGAAATACTGTGGGGGTTCTGGGGAATCTTTC	6501
QY	6010	tgtagagatgtagactgtgcgtcacctccaaacgatlbgagactgtgaaa tggccgactcttc	6069
Db	6502	TGTAGGAGATGAGACTGCTGCCTCCACCTCAACGATGAGACTGGAATGGCCGATCTCTGC	6561
QY	6070	cagagccgggttaaatcttcctcttgcgttgcattccgcgcggcggaatlaacaaatgtagaa	6129
Db	6562	CAGAGCCGGGTAAATTTCTCTTTGTCAATTCCGGCCGGGCAAGTACAAAGTGTAGAA	6621
QY	6130	ctatataagataagagctgtcctcccatagacttagatatactataagataagtttaact	6189
Db	6622	CTATATAAGATAAGACTGCTGCCATCGACTGAGCTTATGATTACTTAAGATAAGTTTCACT	6681
QY	6190	g - -aaagaaacagggaaataltcoaaggtgataaaltgta tlgactgtgggaatagt	6247
Db	6682	GAAAAAGGAAACAGGAAATATTTCAAAAGGTGATTAAGTATGAGCTGGGGAACTACTT	6741
QY	6248	ttttat-tatatgcyggggaggaaggtccactttaacattcgcttagatagaagc	6306
Db	6742	TTTTATTAATTTGGCGGGGAGGAGGCTCCACTTATTAACCTTGGCTTTAGGATAAGACG	6801
QY	6307	ggagacagaacccccctgtggaacttgggaaccggataaagtaactgtgc tgaacaagggccccc	6366
Db	6802	GGGACAGAACCCCTGTGGGAGTGGGAGCCCGAATAAGTACTGTGACAGGGGGCCCCG	6861
QY	6367	ggcctggaagccacgcaataacttgcggtggtccccaattaacctgcgtgcggtcgaata	6426
Db	6862	GCCCTGGAGCCACCGCATTAATCTGCGGTGCCCAATTAACCTGCTGGGCTGACTACATA	6921
QY	6427	acaaagccgcctlaagcaacagttaccatgtgatgtattcctcaacaagccctagaanaacc	6486
Db	6922	ACAAAGCCGCTTAGCANACAGTACCACCTGATTGATTTCTTACCAACAACGCTAGAAATCC	6981
QY	6487	ccagtggtctccgtgtlaagacaaggaacagagacttcctagctcatccaaaggagcttccaa	6546
Db	6982	CCAGGTGCTCCTGTTAGACAGAGACAGAGACTTCTAGTCTCATCCAGGAGACTTCCAA	7041
QY	6547	gacataactccacggacccctgtatgcacttctctctgtgtgttgcattgttgcataccctcagg	6606
Db	7042	GCAATCAACTCCACGACCTGTATGCACTTCTTCTTGTGGCTTTGTCTATCTCAAGG	7101

QY	6607	ccccctattatctgaggagatcggcctaaagaaagaatltcaatgltgacaaagagataga	6666
Db	7102	ccctccctattatgagggagatggcctaaagaaagaatltcaatgltgacaaagagataga	7161
QY	6667	aatcaatgtacatgagggtcccgaaataaacttaacctcaactgaadtcttcogsgaaggg	6736
Db	7162	AATCAATGTACATGGGGGTCCCGAATAAAGCTTAACCTTCACTGAAGTTCCGGGAAGGG	7221
QY	6727	acatgtacaggaaaagctcccccacccacccaacaccttgcctatgaactcgtglttat	6786
Db	7222	ACATGATAGGAAAAGCTCCCCCAACCCACCAACACCTTTGCAATAGTACTGTGTTAT	7281
QY	6787	gagcaagcctcagaataactcagattttaagctcgtgtataacaagtggtggcagat	6846
Db	7282	GAGCAGGGCTTCAGAAATACGATATTAGTACCTGGTTTAAACAGGTGGGGCAGCAAT	7341
QY	6847	actgagtttaacccctcgtgttttcacctaagcttcaaccaatccaagaatttcgtgtc	6906
Db	7342	ACTGGGTTAACCCCTGTGTTTCACCTAGTCTTAAACCAATCCAAAGATTTCTGTGC	7401
QY	6907	atggtccaaatcgtcccccgagtgtaatacactccctgaggaagtggtccctgtatgaat	6966
Db	7402	ATGGTCCAAATCGTCCCCCAGTGTAATACATCGTAGAAGAGTGGTCTGTATGAATAT	7461
QY	6967	gactatcggtataaacgcagcaaaagaagaaacccgtatcccttaacctagctgtatgctc	7026
Db	7462	GACTATCGGTATTAACCCAGCAAAAGAGAACCGGTATCCCTTAACCTTAGCTATAGTGC	7521
QY	7027	ggatttaggagcggcgtgtgctgtagaacaaggaagcgcgtccctgatacaacggaccag	7086
Db	7522	GGATTAGGAGCGCGGTGTGGCGTAGAGAACGGGACAGCTCCCTATATCAACAGACACAG	7581
QY	7087	cagctagagaaaggaactctgtgtgacatactatgcgcacatcacagaagatccgagccta	7146
Db	7582	CAGCTAGAGAAAGGACTTGGTGAAGTACATACGCGCCATACAGAAAGATTCGCAACCTTA	7641
QY	7147	aaggaatctgttagcaaacctagaagagtcctgaactcttgcgtgaagtggtctacag	7206
Db	7642	GAGGAGTCTGTATTAACAACCTTAGAAGATCCCTTACTTTGCTGAAGTGGTCTTACAG	7701
QY	7207	aaccggaggagatagatctcgtgttcttaaggaaggtgggttatgtgcagcctataa	7266
Db	7702	AACCGAGGGATTAGATCTGCTGTTCTTAAGGAAGGTGGGTATGTGCACCTTTAA	7761
QY	7267	gaagaatgttccctactgtatgaatcaactaagggacatcaaggaactccaatgaacgtc	7326
Db	7762	GAAAGATGTTCCTTCTATGTAGATCACTCAGGAGCATAGAGACTCCATGAGCAAGCTT	7821
QY	7327	agaaaaagtltagagagcgtctgaagggaaagaagagcctgaacagggtgltgtlaagga	7386
Db	7822	AGAGAAAGGTTAGAGAGGGCTCGAAGGGAAAGAGAGCTGACAGGGGTGGTTGAAGA	7881
QY	7387	tggttcaacaggtcctctgtatgataccaacctgtcttctgcctgaagggccctagta	7446
Db	7882	TGTTTCAACAGAGTCTCCTTGATGACACCTCTTTCGTGCTGTACAGGGGCCCTTAGTA	7941
QY	7447	gtccgcgtccgttaacttaagaatgttgagcccttgcttaataatagtttgttgcccttgt	7506
Db	7942	GTCCTGCTCTGTTACTTAAAGTTGGGCTTGGCTTAATTAATAGGTTGTGTCTTGT	8001
QY	7507	agagaacgagtgagtgatcagtcacgaatcaltgltacttaagcaaacgtataccaagcctctg	7566
Db	8002	AGAGACGAGTAGTGTGCAAGTCAGATCATGTGTAAGCAACAGTACCAAGCCTTCTG	8061
QY	7567	agccaaagagaactgaactctatagccttccaggtcttaagatltgaactataaacaagc	7626
Db	8062	AGCCAAAGAGAACTGACCTCTAGCCTTCCCACTTTTAAAGATTAAACATTAACAAAC	8121
QY	7627	aagaagtgagggaatgaagaatgaanaatgaacactaacctccgagaacccaggaagtta	7686
Db	8122	AAGAAGTGGGGAATGAAGAATGAATAATGCAACCTTAACCTTCCCAAGAACCGAAGTTA	8181
QY	7687	ataaaagctctaaatgcccccgaaatctacagacctgtgcgtgctgcagtaaatagtaga	7746

|||||
 Db 8182 ATAAAGCTCTAATGCCCCGGAATTCAGACCCCTGCTGCTCCACATAATAGGTAGA 8241
 Qy 7747 aggtcaacttccatctgtccagggccgtctatctccgtgaagtaagataacaggaaa 7806
 Db 8242 AGGTACACTTGTCTATTCTTCCAGGCGCTGCTATCTGCGCTAAGTAAGATAACAGGAAA 8301
 Qy 7807 tgaattgactaatcgctatctgattctgtctaaactgactgagacacataagaatctga 7866
 Db 8302 TGAGTGTGACTAATCGCTATCTGATTCTGTAAGACGACGACCATAGAGAAATTTGA 8361
 Qy 7867 ttacattgacagccctagctactct 7896
 Db 8362 TTACACATTGACAGCCCTAGTACTACT 8391
 RESULT 6
 LOCUS PEN133818 8750 bp DNA VRL 14-APR-2000
 DEFINITION Porcine endogenous retrovirus type C proviral gag, pol and env genes and LTR (class B, clone 43).
 ACCESSION AJ133818
 VERSION AJ133818.1 GI:6688949
 KEYWORDS env gene; env protein; gag gene; gag protein; pol gene; pol protein.
 SOURCE porcine endogenous retrovirus.
 ORGANISM Porcine endogenous retrovirus
 Viruses; Retrovirdae; Mammalian type C
 retroviruses; 1-Mammalian type C virus group.
 1 (bases 1 to 8750)
 Caudeana, F., Fischer, N., Boller, K., Kurth, R. and Tonjes, R.R. Establishment and characterization of molecular clones of porcine endogenous retroviruses replicating on human cells
 J. Virol. 74 (9), 4028-4038 (2000)
 MEDLINE 20219394
 REFERENCE 2 (bases 1 to 8750)
 AUTHORS Tonjes, R.R.
 TITLE Direct Submission
 JOURNAL Submitted (04-MAY-1999) Toenjes R.R., Medical Biotechnology, Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, Hessen, D-63225 Langen, GERMANY
 FEATURES
 source location/Qualifiers
 1..8750
 /organism="porcine endogenous retrovirus"
 /strain="type C"
 /db_xref="taxon:61673"
 /clone="43"
 1..629
 LTR
 misc_feature
 467
 /note="cap site"
 632..649
 /note="g1y4"
 1064..2635
 /gene="gag"
 1064..2635
 /gene="gag"
 /function="retroviral capsid"
 /product="Gag protein"
 join(1064..2635,2639..6223,6099..8072)
 /note="viral polypeptide"
 /codon_start=-1
 /protein_id="CAB5341.1"
 /db_xref="GI:6688950"
 /translation="MGQVTTPLSLTLDHMTVEVSRANHLVYKKGPMTQTCASMP
 TFDVGMPSSESTSEILIAVKALIFQTSSSHPDDEPIITWQDLAEDPAPWKKPMLN
 KPRKGPRIALCEKKNHSAKEVPSRIYPERIEPTPEPQPVPPYPAQGNRG
 PSAPCPAVVEGCPAAGTSSRGAETPERTEIALPLRTYGPVPPGQQLPQYWPSS
 ADLYNMKTNHPSESDPQRLTGLVESLMSHOPWDCQQLQTLTTEBEREILIEA
 RKNVPADGRPTQLEIDMGFPLTRPGMDYNTAEKRESLKIYRQALVAGASARP
 TNLAKYEVMOGNEPDSVLEKLEMAFRFPPFDSSEKASVALATIGOSALDIR
 KTIORLEIOEALRLDLVRAEKVYVRRTEBEKEDRKRERERERERERERERERER
 KILAAVEEGSSRERERDRKIRSGRSGNIGNRTPLDKDCAVCYCKEKGHARNCPK
 CDS
 mat_peptide
 gene
 primer_bind
 misc_feature
 LTR
 misc_feature
 467
 /note="cap site"
 632..649
 /note="g1y4"
 1064..2635
 /gene="gag"
 1064..2635
 /gene="gag"
 /function="retroviral capsid"
 /product="Gag protein"
 join(1064..2635,2639..6223,6099..8072)
 /note="viral polypeptide"
 /codon_start=-1
 /protein_id="CAB5341.1"
 /db_xref="GI:6688950"
 /translation="MGQVTTPLSLTLDHMTVEVSRANHLVYKKGPMTQTCASMP
 TFDVGMPSSESTSEILIAVKALIFQTSSSHPDDEPIITWQDLAEDPAPWKKPMLN
 KPRKGPRIALCEKKNHSAKEVPSRIYPERIEPTPEPQPVPPYPAQGNRG
 PSAPCPAVVEGCPAAGTSSRGAETPERTEIALPLRTYGPVPPGQQLPQYWPSS
 ADLYNMKTNHPSESDPQRLTGLVESLMSHOPWDCQQLQTLTTEBEREILIEA
 RKNVPADGRPTQLEIDMGFPLTRPGMDYNTAEKRESLKIYRQALVAGASARP
 TNLAKYEVMOGNEPDSVLEKLEMAFRFPPFDSSEKASVALATIGOSALDIR
 KTIORLEIOEALRLDLVRAEKVYVRRTEBEKEDRKRERERERERERERERERER
 KILAAVEEGSSRERERDRKIRSGRSGNIGNRTPLDKDCAVCYCKEKGHARNCPK

KKNKGKVLAEEDKGRGSDPLPEPRVALKEGQVPEFLVDGTGAHSLVLLQPLGK
 KKKSSVMKATQROY PMTTRTVDLAGRVHSLVLPEDPVLPLGLDILTKMAQI
 SEFCGRPEVSNKRPITVTLTDLDEEYVSPVQPDODIOSMLEQFOAAMEQI
 LKQPOVIOAKASATPVSVQYLSSEARGIIPWPHOIQOGLIVPOSPMNTPL
 LPRKPGTNDPVDLREVNKRVDIHPEVNPVNLISALIPERENMTVDIDAF
 CIRLHPTSPPIFAFMRDPTGRTQTLWTRLPQGFKSPITIFDALRLDANRTOH
 PVTLLQIVYDILLAGATRKQDLESTKALLLESDLGASAKKAQICRRELVGLYS
 LGGQRMVLEADKARKTVQIIPAPTTAKQVRELGTAFCRLMIPGFTLAALVPLTK
 KGEFMAPEHOKTFDAIKKALLSAPALPLVTEKPTLVDERKAVAGVLTQGLP
 RRPVAVLSKRLDPAVSGMVCILKAAVALVKKDRLTLCOTVIAVPALENTVRO
 PDRMTNARMTHVOSILLTERVTFAPPAALNPILLPEEDPYTHCHOLLIEG
 VKDLDIPLTEVEVLTPTDSSVYVEKRNAGAAYVDGHTINASSLPBSTSAKAE
 LKALQALRLAEGKSNITTDSTRVFAIYAHVGAITYKORGLTSAKRIKKEETLSL
 LEALHLPKRLAIHCGHQKADLISRONOMADRYAKAQAVALNPITETPKAPBR
 ROYLTEDMOEIKKIDQFSETEGTCYTSYKEILIPKGELEYVOOIHETLGLTKHLO
 OLVRTSPYHVLRLPGVADSVYKHCYCOLVANAISRYVPGKRLSGKHADEVETE
 VKPARKGNKYLILVDFDTSQWVEAYPTKERTSVYAKILIEIFPRGIPVYSDNG
 PAFVAVOSGLAKILIGDMLKLCATRPSSQVEKMTIKETLTKLTETGIDWIA
 LUPFLVFRNTPGQGLTPYELLTGPPPLVEIASVADVLISQPLFSLKLEWY
 RORAMKOLREAYSGEGDQVPHRQVGVSVVRRRAAGLNRKGPVLYLTPTAV
 KVEGISTWHAHVKLAPDPDSGWRGAETENPLKRLHRLVPSNNNSPGOMHPTLS
 RLPLTRGGEPRKRLRPLSFASIAFELTLTIPQASRLIDSNDHRSLSTMILIDP
 DGTAVNSTRGVAPRGTMPELHPLRLINDAVKSTPPLNLSYGFYCCPGTEKRC
 GSESEPCRRMSCVTSNDGKMPISLDRKFKSTVNGCPKRYVMKLYKPKSSPSD
 LDYLTSTFEKQENIOWIMGSMWGVFTKYGAGASTILRLRLTETPEVAVG
 PDKVLAEGQPPALPEPHNLVPQTLSLPDLTPQPSNGTTLIPNTPRNSPVKPT
 GORLELSIOGAFOAINSDPADTSCMCLDSGPEYGMKKEGFVNTKRNQCT
 GSRNKLITTEVSGKTCIGKAPPSHOHLCTSVYQEAEMQYLVPCYRNMAKMTGL
 TPCVSSVFNQKDCVMVQIAPRVYVPEEVLVEYQVSEMYRNPAREVSLTAVWGL
 LGTAVGVGTALITGPOLNPELHPLRLINDAVKSTPPLNLSYGFYCCPGTEKRC
 QNRRLDILFLREGGLCAALKECCFVYDHSAGATIDSRSKRLERERRERREADQCH
 FKGWNRSPMTWTLTSLALGPLVLLLTLLTLPCLINRFVAFVREVSAYOIMVLRQ
 YQGLLSQGETD"
 2639..6223
 /function="pol"
 /gene="pol"
 /product="pol protein"
 2639..6223
 /gene="pol"
 /gene="pol"
 6099..8069
 /gene="env"
 /function="retroviral envelope"
 /product="Env protein"
 6099..8069
 /gene="env"
 8122..8750
 LTR
 8643..8648
 BASE COUNT 2288 a 2214 c 2286 g 1962 t
 ORIGIN
 Query Match 55.2%; Score 4532; DB 14; Length 8750;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 7862; Conservative 0; Mismatches 51; Indels 8; Gaps 5;
 Qy 276 cgcgccttcggttcgactgtctctcagacgtaagactgagagctgtatcaaga 335
 Db 752 CGCGCTTGTGGTTCGAGCTGCTCTCAGACCTGAAGACGAGAGACTGTGATCAGCA 811
 Qy 336 gacgttctaggagatcaagactgacacccgggggagagcccgagggttgaggagac 395
 Db 812 GACGTCTAGAGAGATCAGAGCTGACCTGGGGAGAGCCCGGAGGTGGAGAGC 871
 Qy 396 caaggagcctgtgtgtccctacttcgltcagaaggagacggttcgttgtaagcga 455
 Db 872 CAGGAGCGCTGTGTGTCTCTACTGTGCTGAGAGACCGAGTCTGTGTTGAAGCGA 931
 Qy 456 aagcttccctcccgagcgcgtccgactcttgctcgtgttggaagacgagcgggt 515
 Db 932 AACCTTCCCTCCCGGCGGCGGCTGCTTGTTCCTGCTTGGAAGACGCGGAGGT 991
 Qy 516 cgcgtgtcgtgactgtgttcctcgtcgtcgtgtcgtcgtcgtcgtcgtcgtcgtcgt 575
 |||

QY 2735 aattggccgcatgttcaagaatlaaccacaaggacatcctagttccctgttccaaatccc 2794
|||||
Db 3211 AATTGGCGCGCATGTTCAAGATTAATCAACAGGGGCACTGATCTGTCATATCCC 3270
QY 2795 ttggatactccctgtcaccggttagaagaccttggagccaatgtattatgcaagtaca 2854
|||||
Db 3271 TTGGATACTCCCTGCTACCGGTTAGGAAGCTGGGACCAATGATATTCACACAGTACA 3330
QY 2855 ggaattgagaagtgtaataaagggtgcaagacatacaccacaagctccgcaacctta 2914
|||||
Db 3331 GGACTTGAGAGAGGTCAATTAAGAGGTGACAGACATACACCAAGGTCCGAGACCTTA 3390
QY 2915 taaccttggagcgccctcccgctgaaaggaactgtatgcaaglatatgaaataaaga 2974
|||||
Db 3391 TAACCTTTGAGCGCCCTCCCGCTGAGACGAGACTGTACACAGTATGTGACTTAAAGA 3450
QY 2975 tgccttctcgtcgaatatacaccaccacagcccaaccttcttccttcgaaatgag 3034
|||||
Db 3451 TGCTTTCTTTCGCTGAGATTAACACCCACACGCAACCGCTTTTGGCTTCGAAATGAG 3510
QY 3035 agatccaaagtacgggaagaacggcgagctcaacttggaccgacttgcacaaagtaca 3094
|||||
Db 3511 AGATCCAGGTACGGGAAGAACCGGGGAGCTCAGCTGAGCCGAGCTGCCCAAGGTTCA 3570
QY 3095 gaactcccgacacatctttagcagaagccctacacagagacttggccaactcaagatcca 3154
|||||
Db 3571 GAATCCCGGACCAATCTTTGAGAGGCTTACACAGGAGCTTGCCAACTTCAGATCA 3630
QY 3155 aacacctcaagtgacacclctcctcagtagtgatgacatgcttcttggcgggaagccaca 3214
|||||
Db 3631 ACACCTCAGGTGACCCCTCCCTCCAGTACGTGATGATGACTGCTTGGCGGGAGCCACCA 3690
QY 3215 acaagactgcttagaagatgacgaagacacttgcgtgaaattgtctgaacttagctacag 3274
|||||
Db 3691 ACAGGACTGCTTAAGAAAGTACGAAGCACTACTGTAATGTGCTGACCTAGGCTACAG 3750
QY 3275 agccttctgaagaagcccaagatttgcagagaagtgtaactacttgggtgaagttc 3334
|||||
Db 3751 AGCCTCTGCTTAGAGGCCCAAGATTGTCAGAGAGAGATTAACACTTGGGGTACAGTTT 3810
QY 3335 gcgggcgaggcagatgctgacggaagcagcgaagaagaactgtatccagaatacgcgc 3394
|||||
Db 3811 GCGGGGCGGGCAGGATGCTGACGAGGCGACGGAAGAAACGTGATGCCAGTACCAGC 3870
QY 3395 cccaacacagcccaaaagttagaagatttttggggagacgtgtgatttgcagactgtg 3454
|||||
Db 3871 CCAACCCACAGCCAAACAAATGAGAGATTGTTGGGACAGCTGATTTGACAGACTGTC 3930
QY 3455 gatcccgagggttggagaccttagaccccaacttaccgcttacaacaaagaagaaggcg 3514
|||||
Db 3931 GATCCCGGGGTTTGGCAACTTAGACCCACCTCTACCCGCTTACCAAAAGAAAAGGGA 3990
QY 3515 attctcctgggtctctgagcaccagaagcaattgtatglatcaaaaagccctgtcag 3574
|||||
Db 3991 ATTCTCTGCGGCTCTGAGACACAGAACATTTGATGCTATCAAAAAGCCCTGCTGAG 4050
QY 3575 cgaacctgctctggccctccctgaagtaactaaaccttacccttattgtgtatgagcg 3634
|||||
Db 4051 CGGACCTGCTCTGGCCCTCCCTGACGTAACTAACCCCTTATGATGATGAGCG 4110
QY 3635 taagggaatagcccgagagatttaacccaacccctagagacatagagaagccgtgtgc 3694
|||||
Db 4111 TAAGGAGTAGCCCGAGAGATTTAACCAAAACCTTAGACCAATGAGAGAGACTGTTGC 4170
QY 3695 ctactgtcaagaagactgtatcctgttagccagtggttggccgtatgtctaaagctat 3754
|||||
Db 4171 CTACCTGTCAAAAGAACTGTATGCTGACAGTGTGTTGCCGATATGCTTAAGGCTAT 4230
QY 3755 cgcagctgtggccatctgttcaagaagcgttgaacaaattgacttgggacagaatataac 3814
|||||
Db 4231 CGCAGCTGTGGCATTAAGGTCAAGAGACGTGACAAATTTGACTTGGGACACAAATATAC 4290
QY 3815 tgaatagaccccatgcatltagagaaatcgttgcgacgcccagaccgaltgatgac 3874

Db 4291 TGTATATGCCCCCATGATTTGAGAAATCTGTTCCGACACCCCAAGCCAGATGATGAC 4350
QY 3875 caacgccgcatgacccactatacaagccgtcttctacagaagaggttcaacttgcctgc 3934
|||||
Db 4351 CAACGCCCGCATCACCCACTATCAAAAGCCGTCTTCTACAGAGAGGCTGACTTGGCTGC 4410
QY 3935 accagcgctctaacctctccactcttctgcttgaagagactgtatgaacagttgactca 3994
|||||
Db 4411 ACCAGCGCTCTAACCTCTCCACTCTTCTGCTGGAAGAGACTGATGAACCACTGACTCA 4470
QY 3995 tgaatgcatcaactatgtatgagagagcttgggtccgcaagagacttacaataacc 4054
|||||
Db 4471 TGATTTGCATCACTATTTGATGAGAGACTGGGGTCCCAAGAGACTTACACACTTACC 4530
QY 4055 gctgactggagaagtctcaactcgttctcaactgaaggaagcactatgtgtggaagtaa 4114
|||||
Db 4531 GCTGACTGGAAGAGTCTTAACCTGTTCACTGACGGAACAGCTATGTGTGAAGTAA 4590
QY 4115 gaagatggtcgtggcgagtggtgagcggagccgacgacttggcgcaagcgtgc 4174
|||||
Db 4591 GAGGATGTGCTGGGGCGGGGTGTTGACCGGACCCACAGATCTGGCCAGACCTGTC 4650
QY 4175 ggaaggaactcagcgcaaaagcgtgagctcatggtccctcaagcaagcttgcgctgc 4234
|||||
Db 4651 GGAAGGAACCTTACGSCAAAAGGCTGAGCTCATGGCCCTCAGCAAGCTTTCGGCTGCG 4710
QY 4235 cgaaggaatcatataaataattatagcgaagcagatgactccttgcagctgcacgt 4294
|||||
Db 4711 CGAAGGGAATCCATTAACATTTATACGACAGCAGATGCTTGGCAGCTACAGCT 4770
QY 4295 acagggcgcatctataaacaagaagggtgtgttacccttcaagaaggaggaataaaga 4354
|||||
Db 4771 ACAGGGGGCATCTATAAACAAGGGGTTGCTTACCTCAGCAGGAGGGAATATAAGA 4830
QY 4355 caaagsgaaattctaagcctatagaaagccttaccatttgcacaaaagcgtactatcat 4414
|||||
Db 4831 CAAAGGGAATTTCTAAGCTTATAGCCTATTAGAACCTTATGCCAAAAGGCTATGATAT 4890
QY 4415 acactgtctgacatcagaagaagcccaagatctatctatagagggaacagatgtcga 4474
|||||
Db 4891 ACAGTGTCTTGACATCAAGAAAGCCAAAGATCTCATCTATGAGGGAACCGATGGCTGA 4950
QY 4475 ccgggttgcgaagcagcgccagcgtttaaactcttgcctataatagaaagcccaa 4534
|||||
Db 4951 CCGGGTGGCAAGCAGGAGGCCAGCGTAACTTGTGCTTATAATAGAAAGCCCAA 5010
QY 4535 agccccaagaaaccagaagcagtlacacccctagaagacttgcgaagataaagaataga 4594
|||||
Db 5011 ACCCCCAAGCAACACAGACAGTACACCTAGAGACTGCAAGATTAATAAAGATAGA 5070
QY 4595 ccagttctctgagactcggaggagacgtgtatatacccatataggaagaaatcccgcc 4654
|||||
Db 5071 CGAGTTCTCTGAGACTCCGGAGGGGACTGCTATACCTCATATGAGGAAGAAATCTGCC 5130
QY 4655 ccacaagaaggttagaatatgtccaacagatacatcgtctaaaccaactaggaactaa 4714
|||||
Db 5131 CCACAAGAAGGGTTAGATATGTGCCAACAATATCATGTCTCACCCACTAGAGACTTA 5190
QY 4715 aaactgtgaagatgtgtcagaacaatcccttcatatgttctgaagcttaccagagtgyc 4774
|||||
Db 5191 ACACCTGACAGCATTTGGTCAGAACATCCCTTATCATGTTCTGAGCTACAGAGTGGC 5250
QY 4775 tgaactggttgcataaactgtgtgacctgacagctgttgaatgtataccttccagaat 4834
|||||
Db 5251 TGACTCGGTGTCAAACATTTGTGTGCTGCACTGTTAATGCTATATCTTCCAGAGT 5310
QY 4835 aactccagaaagagactaaagggaagccaccccgagcgtcactgtggaagtgaactcac 4894
|||||
Db 5311 ACTTCAGAGAAAGAGACTTAAGGGAAGCAACCCAGGGGCTACCTGGGAAGTGACTTAC 5370
QY 4895 tgaagtaagccggtctaatacgaagaacaatatctatgtgttctgttagaaccttctc 4954
|||||

Db 5371 TGAAGTAAGCCGGCTAAATACGAAATAATCTATGTTTGTAGACACTTTTC 5430
Qy 4955 agagatgttagagcttatctactaagaagaagacttcaacggtgtgttaagaat 5014
Db 5431 AGGATGGGTAGAGCTTATCTACTAAGAAAGAGACTCAACCGTGTGCTAAAAAAT 5490
Qy 5015 actggaagaatttttccaagattgnaactaagatagataggttcacagaaatgtgc 5074
Db 5491 ACTGAGAGAAATTTTCCGAGATTGTGGAATACCTAAGATATGCGGTCAACAAATGTGTC 5550
Qy 5075 agcttgcgtcccaagtaagtaagagagctgcccagaataatgtggatgtatgaaact 5134
Db 5551 AGCTTTGTGTCCAGTAGTACAGGAGCTGCCAAGATATGCGATTGATGGAACCT 5610
Qy 5135 gcaattgtcatcacagaccacaaagctcagagacggttagagagaaatagaacctta 5194
Db 5611 GCATGTGCTATACAGACCCCAAGCTCAGACAGTGAAGATGAATAAACCAATTA 5670
Qy 5195 agagacccttaaccaaatltgacacagagactgtgcatatgtatgtgctcctgc 5254
Db 5671 AGAAGCCCTTACCAATTTGACACAGAGACTGGCATTAAATGATGGATACCTCCTGCC 5730
Qy 5255 cttgtgtcttttaggttgaagaaaccccttgacagtttggtgtgacccctagaatt 5314
Db 5731 CTTGTGCTTTTAAAGGTTAGGAACACCCCTGACAGTTTGCGCTGACCCCTATGAATT 5790
Qy 5315 gctctagggggacccccccgtgtgcaagaatgtccttgcacatagtgctgtgt 5374
Db 5791 GCTCTAGGGGGACCCCCCGTGTGTAAGTAATGCTCTGTATCATAGTCTGATGTGCT 5850
Qy 5375 gcttcccaagcttgtctccttagctcaagcgctcaggtggtgtgagcagcagtg 5434
Db 5851 GCTTCCACAGCTCTGTCTCTAGGCTCAAGGCGCTCGATGGGTGAGCAAGACGCTG 5910
Qy 5435 gaagcagctccgggagagcctactcaag -- aggaactgtgcaagttccacatgctcca 5491
Db 5911 GAAGCAGCTCCGGAGGCTCTACTACAGAGAGAGAGATTGCAAGTTGCCATCTGTTCCA 5970
Qy 5492 agttgagatctgactatgttagacgacacggtgagagaacctgaacctgtgtgga 5551
Db 5971 AGTGGGAGATTCACTATATTTAGACCCACCCCTGCAAGAACCTCGAGCTGGTGGAG 6030
Qy 5552 gggacactatctgacttltgacacacacacagcgctgtgaaagtgaagaatccccac 5611
Db 6031 GGGCCCTTATCTGCTACTTTTGAACACACACACGCGGTGAAGTGAAGATCTCAC 6090
Qy 5612 ctggtatccatgcatcccaagtttaagcgcgcacactcccgactcgggtgtgagagcga 5671
Db 6091 CTGATCTCATGTCATCCACGTTAAGCTGGCGCACCTCCGACTCGGGGGTGGAGCCGA 6150
Qy 5672 aaagactgagaatcccttaagcttgccctcaatgcccgtgtcctactactaacaata 5731
Db 6151 AAAGACTGAGATCCCTTAAGCTTGCCCTCCATCGCTGTCTTACTTCAACATAA 6210
Qy 5732 ctcccagcgcagtagtaaacgcttataagacgctcgaacccccccatagaccttacc 5791
Db 6211 CTCGCCAGGCGCAGTAGTAAGACGCTTATAGACGCTCGAACCCCATAGACCTTTAACC 6270
Qy 5792 ttaactgctgattatagaccctgatacgggtgtcaactgtaataagactcgaagtggtg 5851
Db 6271 TTACTCTGGTGAATTAATGACCCGTGATACGGGTGTCACTGTAAATACACACTCGAGGGTGG 6330
Qy 5852 ctcttagagcactgtgtgctgaactgtatcttgcctccgattgataaccccgctg 5911
Db 6331 CTCTTAAGGACCTGTGTGCTGACTGCACTTTTGTGCTCCGATTGATTAACCCCGCTG 6390
Qy 5912 ttaaaagcacactcccaactagtcgtagtatagtgtctatgtctgcccagagcag 5971
Db 6391 TTAAGACACACCTCCCAACTAGTCCGTAGTATGAGTCTATGCTGTGCCAGGACAG 6450
Qy 5972 agaaagagaataactgtgggggtcttggggaatccctctgtlaagagaatgagctgtgca 6031
Db 6451 AGAAAGGAATACTGTGGGGTCTGTGAAGAACTCTTGTGTAGAGATGAGAGCTGGCTCA 6510

Qy 6032 cctccaacgataagagacttgaaatgtgcccagatcctctcccaagaccggglaaaattcct 6091
Db 6511 CCTCCACAGATGAGAGCTGGAATATGGCCGATCTCTCTCCAGACCGGGTAAATTTCTCT 6570
Qy 6092 ttgtcaattccggccggcgaagtacaaatgatgtaaactalatalaagaataagactgtct 6151
Db 6571 TTGTCAATTCGGCCCGGCGCAGGTACAAAGTATGATGAACCTATATTAAGATAAGACTGCT 6630
Qy 6152 ccccatcgaacttagataltactaaagataagtttcactg -- aaagaaacaggaataa 6205
Db 6631 CCCCATCAACCTTACATATCTAATGAAGATTAAGTTTCACTGAAAGAAAGCAAGAAATA 6690
Qy 6210 tccaagaatgataaaltgtatgaactgtggaactgtttttat - tataitggggggag 6268
Db 6691 TTCAAAAGTGAATTAATGCTATGACTGGGAAATGTTTTTTATTAATATGGCGGGAG 6750
Qy 6269 cagggttcaacttcaacatctgccttaagatagaagacgggagacacccctgtgcaa 6328
Db 6751 CAGGGTCCACTTAAACCATTCGCTTAGATAGAGACGGGACAGAACCCCTGTGGCAG 6810
Qy 6329 tggagccgataaagtaactgtgatacaggggccccggccctggaagccacgataact 6388
Db 6811 TGGGACCCGATTAAGTACTAGCTGGAACAGGGGCCCGCCCTGAGACACCGCATTAAT 6870
Qy 6389 tgcggtgtcccaattaaactcgtcgtgcggccctgacataacagcgcgctagcaacagta 6448
Db 6871 TGCCTGCTCCCAATTAACCTGCTGCGGCTGACATATACAGCCGCTAGCAACGGTA 6930
Qy 6449 ccactgtatgtatctccacacacagcclagaacatccccaggtgtctcgttlaagaag 6508
Db 6931 CCAGTGTGATGATTTCTTACCAACAGCCCTAGAAATCCCGAGTGTCTGTAGACAG 6990
Qy 6509 gacagaagactcttcaagttcactcaagggagcttcccaagccataactccacgcagctg 6568
Db 6991 GACAGAGACTCTTCACTGATCATCAGGAGAGCTTTCCAAACCAATCAACCGACCGCTG 7050
Qy 6569 atgcacactctctcgttgcgttltgacttcaactcaagggccctcattatagaggagag 6628
Db 7051 ATGCCACTCTCTTGT 7110
Qy 6629 ctaagaagaagaatcaatgtgacccaagagcatalagaatcaatgtataggggtccc 6688
Db 7111 CTAAAGAAAGAAATTAATGTGACCAAAAGACATAGAAATCAATGTAGGGGTGCC 7170
Qy 6689 gaataaagttacccttaactgaagtttccgggaagggagactgcaatgaagaagtcccc 6748
Db 7171 GAATATAGCTTACCTCACTGAAGTTTCCGGGAAGGACATGCAATAGGAAAGCTCCCC 7230
Qy 6749 catcccaacaacacttgcataagtaactgttcaagcagcctcagaanaatcag 6808
Db 7231 CATCCACCAACACTTTGCTTACTGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7290
Qy 6809 atttagtactgtgtlataaacaaggtgtgggaatgcaatactgtgttaacccccgtgtt 6868
Db 7291 ATTATGATCTGTTATTAACAGGTGTGGCGCTGCAATACTAGGTTAACCCTCTGTGTT 7350
Qy 6869 ccacactagcttcaacaaatcaacaaagtttctgtgtatgttccaatctgtccccag 6928
Db 7351 CCTCTCACTTCAACCAATCAAGATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7410
Qy 6929 tgaactacatcctgaaggaagtgctcctgtatgaatagatcatgatactgataacagacaa 6988
Db 7411 TGATACATCTCTGAGAGAGTGTGCTTGTGATGATATGATGATGATGATGATGATGATG 7470
Qy 6989 aaagagaacccglatcccttaacactgactgtgaatgtcctggtatgaagcagcgtgtg 7048
Db 7471 AAAGGAACCCCTATCCCTTACCTAGCTGTATCTGTGATTAAGGAGCGCGTGTGGCG 7530
Qy 7049 taggaacggagcagctgcccgtatcaagaagccacagacttagaagaagactgtgtg 7108
Db 7531 TAGGAACAGGACACTGCTGCTGATACAGGACACACAGCTGTAGAGAAAGACTGTGCTG 7590

QY 7109 agctacatgcygcacatgacagaagatctccgagccttaagagatctgttgcgaacctag 7168
 |||||||
 Db 7591 acctacatgcccgcacatgacagaagatctccgagccttaagagatctgttgcgaacctag 7650
 |||||||
 QY 7169 aagaatccctgacctcttgcctgaagtgccttctacagaaccgagagagatctgac 7228
 |||||||
 Db 7651 aagaatccctgacctcttgcctgaagtgccttctacagaaccgagagatctgac 7710
 |||||||
 QY 7229 tgccttctacagaagaagtgagatgctgacgaccttaaaagaagaatgttgccttctatgag 7288
 |||||||
 Db 7711 tgccttctacagaagaagtgagatgctgacgaccttaaaagaagaatgttgccttctatgag 7770
 |||||||
 QY 7289 atcaatcagagagccatcagagatcctacagaagaagcttagaaaaagttagagagagctc 7348
 |||||||
 Db 7771 atcaatcagagagccatcagagatcctacagaagaagcttagaaaaagttagagagagctc 7830
 |||||||
 QY 7349 gaaggaagaagaagctcagagagagtggtttagagaagatgcttcaacagagctcctctgga 7408
 |||||||
 Db 7831 gaaggaagaagaagctcagagagagtggtttagagaagatgcttcaacagagctcctctgga 7890
 |||||||
 QY 7409 tgaccacccctcttctgctctgacgagggccctagtagtctcctctgcttctacag 7468
 |||||||
 Db 7891 tgaccacccctcttctgctctgacgagggccctagtagtctcctctgcttctacag 7950
 |||||||
 QY 7469 ttggagccttgccttaataatagatgttgccttgccttgccttagagaacagagtgagtgacatcc 7528
 |||||||
 Db 7951 ttggagccttgccttaataatagatgttgccttgccttgccttagagaacagagtgagtgacatcc 8010
 |||||||
 QY 7529 agatcatgttacttaggaagaagatcacaagagccttctcagcagaagaagaactgacctct 7588
 |||||||
 Db 8011 agatcatgttacttaggaagaagatcacaagagccttctcagcagaagaagaactgacctct 8070
 |||||||
 QY 7589 agccttcccaagtlcttaagaatlagaaacttaacaagaagaagaatgagggaatgaaagat 7648
 |||||||
 Db 8071 agccttcccaagtlcttaagaatlagaaacttatacaagaagaagaatgagggaatgaaagat 8130
 |||||||
 QY 7649 gaaatgtgaacctaaacctcccgagagccaggaagaatgaaatgaaagcttcaatgtccccc 7708
 |||||||
 Db 8131 gaaatgtgaacctaaacctcccgagagccaggaagaatgaaatgaaagcttcaatgtccccc 8190
 |||||||
 QY 7709 gaattacagaccctgctgctgacgagtaagaatgagtagaagaatgacacttctatgttcc 7768
 |||||||
 Db 8191 gaattacagaccctgctgctgacgagtaagaatgagtagaagaatgacacttctatgttcc 8250
 |||||||
 QY 7769 agggcctgctatcctgagccttaagaagaatacaggaagaatgagtgacttaactcgttact 7828
 |||||||
 Db 8251 agggcctgctatcctgagccttaagaagaatacaggaagaatgagtgacttaactcgttact 8310
 |||||||
 QY 7829 ggattctgtaaaactgactgacacatagagaagaatgattacacatgagacagccctagt 7888
 |||||||
 Db 8311 ggattctgtaaaactgactgacacatagagaagaatgattacacatgagacagccctagt 8370
 |||||||
 QY 7889 acctatcgaactgcaatctgactgacacatgacgagagagccagcagagatgacgacctccg 7948
 |||||||
 Db 8371 acctatcgaactgcaatctgactgacacatgacgagagagccagcagagatgacgacctccg 8430
 |||||||
 QY 7949 agctattttaaaatgattgtgtccacagagcgcgagctctgcgataatttaaaatgattgt 8008
 |||||||
 Db 8431 agctattttaaaatgattgtgtccacagagcgcgagctctgcgataatttaaaatgattgt 8490
 |||||||
 QY 8009 ccagtagagcgagcgctcgcagatattttaaaatgattgtgtgtgagcagacagcgcttgc 8068
 |||||||
 Db 8491 ccagtagagcgagcgctcgcagatattttaaaatgattgtgtgtgagcagacagcgcttgc 8550
 |||||||
 QY 8069 gtgaaccacataaagaagctgtcccgatcccgaccccgagcgccgagctccttaaccctgcg 8128
 |||||||
 Db 8551 gtgaaccacataaagaagctgtcccgatcccgaccccgagcgccgagctccttaaccctgcg 8610
 |||||||
 QY 8129 tgggttagcagctgtgggccccagcgcttgaaataaaatcctctgctgttgcata 8188
 |||||||
 Db 8611 tgggttagcagctgtgggccccagcgcttgaaataaaatcctctgctgttgcata 8670
 |||||||
 QY 8189 a 8189

Db	8671	A	8671
RESULT	7		
LOCUS	PER17013		
DEFINITION	PER17013 7808 bp DNA		
ACCESSION	Y17013		
VERSION	Y17013.1		
KEYWORDS	env gene; gag gene; pol gene.		
SOURCE	porcine endogenous retrovirus.		
ORGANISM	porcine endogenous retrovirus.		
REFERENCE	1 (bases 1 to 7808)		
AUTHORS	Czaderana, F., Fischer, N., Boller, R., Kurth, R., and Toonjes, R. R.		
TITLE	Establishment and characterization of molecular clones of porcine		
REFERENCE	2 (bases 1 to 7808)		
AUTHORS	J. Virol. 74 (9), 4028-4038 (2000)		
TITLE	endogenous retroviruses replicating on human cells		
JOURNAL	20219394		
REFERENCE	2 (bases 1 to 7808)		
AUTHORS	Toonjes, R. R.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-MAR-1998) R. R. Toonjes, Paul-Ehrlich Institut,		
FEATURES	Submitted (30-MAR-1998) R. R. Toonjes, Paul-Ehrlich Institut,		
source	Location/Qualifiers		
	1. 7808		
	/organism="porcine endogenous retrovirus"		
	/db_xref="taxon:61673"		
	/tissue_type="kidney"		
	/cell_line="PK15"		
	260. 1834		
	/gene="gag"		
	260. 1834		
	/gene="gag"		
	/codon_start=1		
	/protein_id="CAA76581.1"		
	/db_xref="GI:3005986"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	7808		
	/db_xref="GI:3005987"		
	/translation="MGQVTPPLSLTLDHMFVRSRAHNLISVQYKKGWQFCASEMP		
	/translation="MGQVTPPLSLTLD		

WRKFFRDLLEYLR"
 5294. -7267
 /gene="env"
 5294. -7267
 /gene="env"
 /codon_start=1
 /protein_id="CAI76583.1"
 /db_xref="GI:3005988"
 /translation="MHPTLSMRHLPTNGEKKRLRLPLSTASIAMLLTLITPQASSK
 RLIDSSNPHRLSLTWLIDPDTGVNSTRGVAPRGTWPELHFLRLINPAVKSTP
 PNLRVSTGFCPPETKEKYCGGSESEFCRWSQVINDDMKMPISLIDRVSEFVN
 SGPQKVKMLYKDKSCSPSDLYLKISTEKGKQENIOMKINMGMSGLFYFYGGGA
 GSTLIIRLITETGPEPVAVGPDVLAEGPPALPEPGLINSDPDATSSCWICISGPPY
 GTTGILPTNPNRSPGVKVGKORLFLIGARQAINSTNDPDRATSSCWICISGPPY
 EGMAKEGFNTRKRNOCNMGSSNRKLTLEVEGSKGICIKAPSHOHLCTIVIEQ
 ASENQYLPGRNWRWACNTGTITPCVSSSVNOSKDFCVMQIVPRVYHPEEVLDEY
 DYRNRRKREPVSLTLAVMLGLTAVGVTGTALITGPQLEKGLGELHAAMTEDLR
 ALSEVSMLSESLTSLSEVYLQNRKIDLPLEGGICALKKECCFYVDHSGAIDS
 MSKLRELRERRRRERADOGMEFEGMFRSPMMTTLISALTGPLVLILLITVGPCLIN
 RFVAFVERVGAQVINVLRQOYGLISGEBTDJ"
 BASE COUNT 2077 a 1971 c 2047 g 1713 t
 ORIGIN

Query Match 49.9%; Score 4096; DB 14; Length 7808;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 7746; Conservative 0; Mismatches 55; Indels 9; Gaps 6;

QY 329 atcagcagacgtgctagagagatcaacagcgtccacccctggggagagcccgaggatg 388
 DB 1 ATCAGCAGACGTGCTAGAGAGATCAACAGCGTCACCCCTGGGGAGAGCGTGG 60
 QY 389 ggaagacccagagacgcgcgtgctcctactgctcaggaagacagcttctgtt 448
 DB 61 GGAGAGCCAGGAGACCGCTGGTGCCTCTACTATGCTCAGAGAGACCGAGTCTGTT 120
 QY 449 gaagagaaagctcccccctccgagcggtccagactcttgcctgcttggagaagcg 508
 DB 121 GAAGCGAAAGCTTCCCTCCCGCGCGCTCGACTTTTGGCTTGGTGAAGAGCGG 180
 QY 509 gacgggtcgtgctgctgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 568
 DB 181 GACGGGTGCTGT 240
 QY 569 cctgtctacagtttaataatgagacagagtaactaccccccctagtttgccttga 628
 DB 241 CCTGTCTACAGTTTAAATATGAGACAGACAGTACACCCCTTAGTTTAACTCTGA 300
 QY 629 ccatttgactgaagttaagaccagagctcaaatltgtcagttcaggttaagaagagac 688
 DB 301 CCATTGGAAGTGAAGTATGATCCAGGGCTCATTAATTGTGCTAGTGAAGTGAAGGAC 360
 QY 689 ttggcagaactctgttcccttgaatgagcaacatcagatgttggatggccatcagaag 748
 DB 361 TTGGCAGACTTCTGTCTGCTCTGATATGCGCAACATTCGATGTGAGTCCATCAAGG 420
 QY 749 gaccttaattctgaatatactcgtgtgttaagcaatcaatlttcagacagagacgg 808
 DB 421 GACCTTAATCTGAATATATCTGGCTGTAAAGCAATCATTTTCAACACTGGACCCAG 480
 QY 809 ctccatccctgacagagacccatatacttctacgttggcagaagtlttgagagaagatcctcc 868
 DB 481 CTCTCATCTGATCAGAGCCCTATATCTTACGTGGCAAGACTTGGCGAGAGATCTCC 540
 QY 869 gccatgtgttaaacatcgtctaaataaaccaagaagcaggtccccgaatcttcttct 928
 DB 541 GCCATGTGTTAAACATGCTTAATATAACCAAGAAAGCCAGGTCCCGGATCTTGCTCT 600
 QY 929 ttgagagaaaaaacaacactcggcgaaagatcgagcccttctcgtatctacccg 988
 DB 601 TTGGGAGAAAAACAACTCGCGCAAAAAGTCGAGCCCTC-TCTGTGTATCTACCCCG 659

QY 989 agatcgagagccgcgcgacttgcgcggaaccccaaccccttcccccaccccttaccag 1048
 DB 660 AGATCGAGAGCCGCCGACTTGGCCGAAACCCCAACCTGTTCCCGACCCCTTATCCAG 719
 QY 1049 caagagtgctgtgaaggaga-cccttgcacctccctggagccctccggttgttgaaggactg 1107
 DB 720 CACAGGGTCTGTGAAGGAGCCCTGCCCCCTCTGAGAGCTCCGTGTGAGGAGACCTG 779
 QY 1108 ctgcgcgagactcgagcccgagagagcgcaaccccggaacgagacagacagatcgcgatat 1167
 DB 780 CTGCCGCGACTCGAGCCGAGAGAGCGCACCCCGGAGGACAGACAGATCCGATAT 839
 QY 1168 taacgtgtgacactatggccctcccatgccaaggggccaattgcaagccctcagatatt 1227
 DB 840 TACCGCTGCGACCATGAGCCCTCCCATGCCGGGGGGCAATTCAGACCCCTCCAGATAT 899
 QY 1228 ggccttctctctcgaatcctataatltgaaacaacacaccccttcccttccgagg 1287
 DB 900 GGCCCTTTCTCTGACATCTCTTAATTGGAAACTACCAATCCCTTCTCGGAGG 959
 QY 1288 atcccacgcctcaggggttgtgtgagctccctatgttctcaccagcctactctgg 1347
 DB 960 ATCCCAACGCTCCAGGGGTTGGTGGAGTCCCTATGTTCTTCACACAGCTACTGG 1019
 QY 1348 atgattgtcaacagctgctgacagacactctcaaacccgagagagagagaaattctgt 1407
 DB 1020 ATGATTGTCAACAGCTGCTGCAAGACACTTTCACAGCCGAGAGACGAGAGATTCGT 1079
 QY 1408 tagagctagaanaaatlttcttggggcgagagcgcaacccagatltgcaaatgaga 1467
 DB 1080 TAGAGCTAGAAAAAATTTCTCTGGGGCGAGGCGCACCCAGCAGATTCGAAATAGAG 1139
 QY 1468 ttgaaatggatctcccttgcactgcgcgcgttgggaactaacaacgcttgaagatagg 1527
 DB 1140 TTGACATGGGATTTCCCTTGACTCGCTCGGTTGGGACTCAACACAGCGGTGAAGATAGG 1199
 QY 1528 agaacttgaataatcctatccagagctctgtgtgagcggtctccgggggctcgaagcg 1587
 DB 1200 AGAGTTTAAATATCTATCGCCAGGCTTGTGTGGGGTCTCCGGGGGCTTCAAGACGGC 1259
 QY 1588 ccaactaatlttgctaaagtaagagagtgatgcagagagacagaaacccctccgttat 1647
 DB 1260 CCAGTAATTTGGCTAAGGTAAAGAGAGTGAAGAGGACCGAGAACCACTCCCTCGGTAT 1319
 QY 1648 ttcttggagagctatgagagccttcaagcggttcaaccccttgcactccagag 1707
 DB 1320 TTCTTGAAGGCTCATGGAAGCCCTTCAAGCGGTTCAACCCCTTGTGATCTTACCTCGAGG 1379
 QY 1708 ccagaaagcctcagtgccctggccttcaatlttggaagctggtcttggatataagaaga 1767
 DB 1380 CCCAGAAAGCCTTCAGTGGCCCTGGCTTCTATTGGGCAAGTGGCTGTGATATCAGAAAG 1439
 QY 1768 aacttcagagacttgaagaggttcaagagagctgaatcgttgcagtgatgagagagcgag 1827
 DB 1440 AACTTCAGAGACTGGAAGGCTTACAGAGGCTGATGATGATGATGATGATGATGATGATG 1499
 QY 1828 agaaagtgtatctcaagaaagagacagaaagagagaaagaaagaaagaaagaa 1887
 DB 1500 AGAAGGGGTATATACGGAAGGAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1559
 QY 1888 gagaagaagaagagaaagcgctgatatagacgcaagaagaagaattgagtaagactctgg 1947
 DB 1560 GAGAGAGAAAGGAGAGAAAGCGATGATAGACGCAAGAGAGAAATTTGACTTAAGATCTTGG 1619
 QY 1948 ccgagtggttgaagagagagcagcagagagagagagagatlttggaaatlaagt 2007
 DB 1620 CCCAGTGTGTTGAAGGAG 1679
 QY 2008 cagagccctgaag 2067
 DB 1680 CAGGCCCTAGACAGTGAAGGAGAGCTGGGCAATAGAGAGAGAGAGAGAGAGAGAGAGAG 1739
 QY 2068 cgtatgttaagaaagagacacttggcagaagagagagagagagagagagagagagag 2127

QY 4288 caacagtacacggggccatctataaacaagggtgtgtctaccctgaagggaggaa 4347
DB 3960 CACACGTACACGGGGCCATCTAAACAAAGGGGGTGTGTACCTCAGCAGGGAGGAA 4019
QY 4348 taagaacaaagggaatcttaagcctattagaagccttaattgcccanaaggctag 4407
DB 4020 TAAAGAACAAAGGAATTTCTAAGCCTATTAGAGCTTACATTTGCCAAAAAGGCTAG 4079
QY 4408 ctattataactgtctctgacatcaagaagccaagaatcatalatctagaagggaacaga 4467
DB 4480 CTATTATACAGTCTCTGACATCAGAACCAAGATCTCATATCTAGAGGGAACAGA 4139
QY 4468 tggctgacgggtgtccaaagcaggccaggctgttaacctctgacctataatgaa 4527
DB 4140 TGGCTGACCGGGTGGCCAAAGCAGCAGCCAGGCTGTAACTTTCGCTTAAATAGAA 4199
QY 4528 cgcccaagagcccaagaccagacagacagacagacccctagaagaactgcaagagataaaa 4587
DB 4200 CGCCCAAGCCCAAGAACCCAGACAGACATACCCCTAGAAAGACTGSCAAGATATAAA 4259
QY 4588 agatagaaccagttctctgagactccggaagggaactgtctatcctaatactgggaagaa 4647
DB 4260 AGATAGACCAATCTCTGAGACTCCGAGGGAGACTGCTATACCTCATATGGAAGGAA 4319
QY 4648 tctctgcccacaaagaagggttagaataatgtccaaacagatacactgctctaacccctag 4707
DB 4320 TCTGCCCCACAAAGAGGGTAAAGATATGTCCAAAGATACATGCTCCTACCTAG 4379
QY 4708 gaactaaacacctgacagactgtgtcagaacatcccttcaatgctctgaagctaacag 4767
DB 4380 GAACATAACCTTCAGCAGAGTGTGTCCAGAACATCCCTTATCATGTTCTAGGGTACAG 4439
QY 4768 gagtgtgactcgggtggtcaacaactgtgtgcccgtccagctgtgttaagttaacct 4827
DB 4440 GAGTGGCTGACTCGGTGGTCAAAACATTGTGTGCTCCAGCTGTTAATGCTAATCCT 4499
QY 4828 ccagaataccctccaggaagaagactaagggaagcccaagggcctcactcgggaatgg 4887
DB 4500 CCAGATATCCCTCCAGAAAGAGACTAAGGGGAAGCCACCGGGCTCCTAGGGAGTGG 4559
QY 4888 acttaactgagtaagacgcgactaatacgaagaacaatactatctgtttttagaaca 4947
DB 4560 ACTTCACTGAGTAAGCGCGGCTAATATGSAACAAATATCTATGTTTGTATACCA 4619
QY 4948 ccttttcaggaatggtagaaggctatacctactaagaagaagactcaccgtgtgtgcta 5007
DB 4620 CCTTTTCAGATGGGTAGAGGCTTATCTACTAGAAAGAGACATTCACCGTGTGCTA 4679
QY 5008 agaaatactgaggaanaatctttccaagatttggaatacctaaggtiaaagtcagaca 5067
DB 4660 A-AAATATCTGAGGAAATTTTTCGAGATTTGGATACCTTAAGGTATTCGGTCCAGACA 4738
QY 5068 atgtccagcttctgttcccaagtaagtcagagactgcccagaatattggggaatgact 5127
DB 4739 ATGTGCCAGCTTTTGTTCACAGTAATCAAGGACGTGGCAAGTATTTGGGATGATT 4798
QY 5128 ggaacatgcatgtgtacatacgaacccaagactcagaacaggttagagaagtagaagaa 5187
DB 4799 GGAACACTGATGTGATACAGACCCCAAGCTCAGGACAGTGAAGATGAATATAA 4858
QY 5188 ccattaaagaaccccttaaccaatgacacagagagactgcatbaatgtgagtgctc 5247
DB 4859 CCATTAAAGAGACCTTACCAATATGACACAGAGACTGCAATTAATATGATAGTTC 4918
QY 5248 tctctgaccttgtgtctttttagagtgaggaacacccctggaacgtttggctgacccct 5307
DB 4919 TCTGCGCTTTGTGCTTTTAAAGTTAGAACACCCCTGAGACAGTTTGGCTGACCCCT 4978
QY 5308 atgaattgtcttaaggggagaccccccgttggcagaatgtgaccttcaatagtgtg 5367
DB 4979 ATGAATTGCTTACGGGGAGCCCGCCGCTGTGTAATAATGCTTCTGTACATATGCTG 5038

QY 5368 atgtgtgcttcccaagccttcttctctgaagctcaagcgctcagtggtgtgaagcagc 5427
DB 5039 ATGTGTGCTTTCCAGACCTCTGTTCTGTAGSCTCAAGGCGCTGAGTGGGAGCAAC 5098
QY 5428 gacgttgaagcagctccggagagccctactaagg--agggaacttgcgaatctcaatc 5484
DB 5099 GAGCGTGAAGAGCTCCGGAGCGCTACTCAGGAGAGGAGACCTTGCAGATTCACATC 5158
QY 5485 gcttccaagtgtgagatctagctatgttagacgcacacgctgcagaagaacctcgagactc 5544
DB 5159 GCTTCCAAATGGAGATTCAGTATGTATGTAGACGCCACCGCTCAGAGAAACCTCGAGACTC 5218
QY 5545 ggtggaaggagaccttaccgtaacttctgacacacccaagcgctgttgaagtgcgaagaa 5604
DB 5219 GGTGGAAGGGCCCTTATCTCGTACTTTGACCACACCAACGCGCTGAAAGTCGAAGGAA 5278
QY 5605 tcccacactgatacctatgatacccaagcttaagcgcgcacacttcccgacttgggttga 5664
DB 5279 TCTCCACTGAGATTCATGCAATCCACAGTTAACTGGCGCCACTCCGACTGGGGTGA 5338
QY 5665 gacggaagaagacttgaaatcccttaagcttgcctccatcgctgttctactata 5724
DB 5339 GAGCCCAAAAGACTGCAATCCCTTAAGCTTGCCCTCATGCGCTGCTTACTCTA 5398
QY 5725 acataactccccaagcagtagtaaaagcctatagacagctcgaaaccccatagact 5784
DB 5399 ACATTAACCTCCAGGCCAGTATTAACGCTTAATAGACAGCTCGAACCCCATAGACT 5458
QY 5785 ttatcccttaactgctgtatatttagacccctgataaggggttaactgttaagtgcactga 5844
DB 5459 TTATCCCTTAACCTGCTGTATTAATGACCCGTAAACGGGTGTCAGTAAATAGCACTGCA 5518
QY 5845 ggtgtgtctccctgaaggaacactgtgtgcctgaactcatttctgctccatattgaac 5904
DB 5519 GGTGTGTCTCTAGAGGCACTGTGTGCTGAACTGCATTTCGCTCCGATTGATTAC 5578
QY 5905 cccgctgttaaaagcaccctcccaactagtcctgaattatggttctatgtctgcca 5964
DB 5579 CCGCGTTTAAAGCAACCTCCCACTAGTCCGTAATGATGGGTATGCTGCCCCA 5638
QY 5965 ggcacagaagaagaatactcgtgtgtgtgtcgtgggaatcctctgtgaggaatgagc 6024
DB 5639 GGCACAGAGAAAGAAATTAATCTGGGGGTTCTGAGGAATCCTTGTGAGGATGAGAC 5698
QY 6025 ttgcgtcaactcccaagatgagacgtgaaatggtcgatctctccagaagccggttaaa 6084
DB 5699 TCGCTCACTCCCAACATGAGACTGGAATGGCCGATCTCTCCAGAGCGGTTAAA 5758
QY 6085 ttctccttltgcaattccggcccggaagtaacaaatgataactataaagataag 6144
DB 5759 TTCTCCTTTGCAATTCGGGCCCGGCAATGACAAAGTGTAAACTATATTAAGATAAG 5818
QY 6145 agctgtctcccatcagaacttagatltactaagaatagttactg--aaaggaacag 6202
DB 5819 AGCTGCTCCCATCAGACTTATGATTAAGATTAAGTTTCACTGAAAAAGCAAAACAG 5878
QY 6203 gaaataatcaaaagtgtataatgtatagctgtgggaatgttttlat-tataatggc 6261
DB 5879 GAAATATTTTAAAAAGTGTGAATATGATAGCTGGGGAATGATTTTATTAATATGAG 5938
QY 6262 gggggagcaggttccacttaaccatcgccttaagatagagaacggggaacacccct 6321
DB 5939 GGGGAGCAGAGGATTCATTAACCATTCGCTTAAAGATTAAGACGGGGAAGAACCCCT 5998
QY 6322 gtggcaatgggagccgaataagtaactgtgtgaacaggggcggccctgagacagc 6381
DB 5999 GTGGCAGTGGGACCCGATTAAGTACGTGTAACAGGGGGCCCCGGCCCTGAGAGCCAGC 6058
QY 6382 cataacttgcggttgcacaaatlaacctgcgtgcgtgacatlaaacaagccgctagc 6441
DB 6059 CATAACTTGCCTGCGCCCAATTAACCTGCTGCGGCTGACATTAACAGCCGCTAGC 6118
QY 6442 aacagtaacacatgatatctactacaaacagcctagaacactccccaagtgctccgtt 6501

Db	6119	AACGGTACACACTGAGTTGATTCCTACCAACACGGCTGAGAAATCTCCAGTGTCTGT	6178
Qy	6502	aagaacagagacaaagaactcttaagctatccagggagcttccaagcatcaactccac	6561
Db	6179	AAAGACAGACAGAGACTCTTAGTCTATCCAGGAGACTTTCCAAACCATCAACAC	6238
Qy	6562	gaacctgatgcactctctctgttggcttgtctatccctcaagggctccattatag	6621
Db	6239	GACCCGTAGTGCACATTCTTTGTTGGCTTTGTCTATCTCAGGGCTCTTATATAG	6298
Qy	6622	gggatggctcaaaagaagaatctaatgtgacaaagaacatagaatcaatgtacatg	6681
Db	6299	GGGATGGCTAAAGAAAGGAAAAATTCATAATGTACCAAGACCATAGAAATCAATGTCAATGG	6358
Qy	6682	gggtcccgaaataaagcttaacctactgaagttccgggaagggagacatgacataga	6741
Db	6359	GGGTCCCGGAAATTAAGCTTAACCTCAGTGAAGTTCCGGGAGGGGACATGCTAGGAAA	6418
Qy	6742	gtctccccaatcccaacaacaccttgcctatagtaactgtgttatatgagcaggtccagaa	6801
Db	6419	GCTCCCCCATCCACCAACACTTTGGCTATAGTACGTGGTTATAGAGAGGCTCAGAA	6478
Qy	6802	aatcagatattagtaacctggctataaagaatgggtggagatgaatcgggtttacccc	6861
Db	6479	AATCAGTATTTTAGTACCTCGTTATACAGAGTGGTGGCATGCAATACGTGGGTTAACCC	6538
Qy	6862	tgtgttccacctgaagctcttaaaccaatccaaagatllctgtgtcatgtgtccaatcgtc	6921
Db	6539	TGTGTTTCTCTCCTCAGTTCCTAACCAATCCAAAGATTTCTGTCTATGTGTCCAAATGCTC	6598
Qy	6922	cccagagtactacacatccctcgagaagatggttccttgaaatatgactatccgtatac	6981
Db	6599	CCCCAGTGTACTACATCCTGAGAGAGTGTGCTTGATGAATATGACTATCGGTATAC	6658
Qy	6982	cgaccaaaagaagaaccgatactccctaacctagctgtatgtctcgataatgaggaagcgc	7041
Db	6659	CGACCAAAAGAGAACCCGTAATCCCTTACCCGTAGTGTATGCTGGATTTAGGGACGGCC	6718
Qy	7042	gttggcgttagaagaagaagatgtgcgcctgatcaagaacacagcagctagagaaga	7101
Db	6719	GTTGGCTGATGAGAACAGGAGACAGCTGCCCTGTATCACAGGACCAACACACTGTAGAAAGGA	6778
Qy	7102	cttgtgtgactacatgtgcgcactgacagagaatctcccgagccttaagaagatctgtgac	7161
Db	6779	CTTGCTGACTACATGTGGGGCATGACAGAAATCTCCAGCTTTAGAGAGTCTGTTAGC	6838
Qy	7162	aacctagaagaatccctgaactcttgtctgaaatgtgttctacagaaccggaaggagata	7221
Db	6839	AACTCAAGAGAGTCCCTGACTCTTTGCTGTGAAGTGTTCTTACAAACCGGAGGGGATTA	6898
Qy	7222	gactgtcgtttcttaagaagaagtggtgttatgtgacgtcttaagaagaatgtgtgttc	7281
Db	6899	GATCTGCGTTCCTCAAGAGAGAGTGGGTTATGTGCAGCCCTTAAAGAGAAATGTTGCTTC	6958
Qy	7282	tatgtagataactcgaagagcatcagagaatccatgaaacaagcttagaaaaagttag	7341
Db	6959	TATGTAGATACATCAGAGGACATCAAGAGTCCATGACCAAGCTTAGAGAAAGTTTAG	7018
Qy	7342	aggagctcgaaggaagaagaaggttgacaggggtgtgttgaagaatgtgttcaeaagttc	7401
Db	7019	AGGGCTGAAAGGGAAGAGAGGCTGACCAAGGGGTGTTTGAAGATGTTTCACAGTCT	7078
Qy	7402	ccctgtagtacaaccctgtcttctgtctgaaggggcccctagtaagctccgtccgttga	7461
Db	7079	CCTTGATGAGACCAACCTCGCTTCTGTCTGACAGGAGACCCCTAGTAGTGTCTCTCTGTTA	7138
Qy	7462	cttaacagtctgggcctgtcttaatatagattgttgccttgtttagagaacgagtgaft	7521
Db	7139	CTTACACTTGGGCTTGTCTTAATATATAGTGTGTTGCTTGTATAGAGAACGAGTGGG	7198
Qy	7522	gcaatcaagaatctgtacttggaaaaagataccaaggtccttgcacaaaggagaact	7581

Dn	7139	GCGTTCACGATCATGTACTTGTTAGGCACAGTACCAGAAGCCCTTCTGAAGCCACAAGAAACT	7258
QY	7582	gaacctagaacgtccccaagttcttaagattlagaactatactaacaagaacaaagatgagggaatg	7641
Dn	7259	GACCTCTTAGCCTTCCTCCAGTTCTTAAGATTAAACAATTATAACAAGCAAGAAGTGGGAATG	7318
QY	7642	aagagatgaanaabycaacacthaaccctcccggaaccgggaagttaataaagaactcaaa	7701
Dn	7319	AAGAAGATGAANAATCAACTTAACCTTCTCCAGAACCCAGGAAGTAAATAAACCCTTAAT	7378
QY	7702	tgcgcccggaatlacagaaccctcgtcgtgcacglaaalagtagtgaagagtcacactcccta	7761
Dn	7379	TGCCCCCGAAATTCACAGACCCGTGCTGGCCAGTAATAGTAGAAGTGTCACACTTCCTA	7438
QY	7762	tttgtccaggagccgtcatctctgcccctaaatgaagataacagagaatgagttgactaatcg	7821
Dn	7439	TGTGTCCAAGGGCCGTGCTACTCTGGCCCTAAATGAATACAGGAATGACTGTCTAATGG	7498
QY	7822	cctaactcggalcttcgtlaaaaactgacbtgcacccalaagaagatgatcacacttgacgc	7881
Dn	7499	CTTATCTIGGANATCTGTAAACTGACTGCGCACCATAGAGAATTTGATTACACATTTGACAC	7558
QY	7882	ccctagtgaaactatctcaactgcgcaatctgltactctcgtcccaagagcccaagacagatgcga	7941
Dn	7559	CCTAAGTACCATCTATCTCAACTGACATCTGTCTACTCTGCCAGAGCCCCACAGATCCGGA	7618
QY	7942	cccccgagactattttaaaatlgattgltccacgagcgcgaggtctcgcataattttaaat	8001
Dn	7619	CCCTCGAGAGCTATTTTAAAATGATGTGTCACAGGACGCGGCTCTCGATATTTTTTAAAAAT	7678
QY	8002	gattgtgtccatbyagcgcgaggtctcgcataattttaaatlgattglttgacgcacag	8061
Dn	7679	GATTGTGTCCACGAGGCCGGCGCTCTGATATTTTAAAATGATGTGCTTGTGACGCGCACAG	7738
QY	8062	cttggtgtbaaccccaataaagctgtcccgattcgcgactcggggccgcaatccctcac	8121
Dn	7739	CTTTGTTGTGAACCCCATAAAGCTGTCCGATTCGCGACTCGAGGCCAGCTCCAGTCTTAC	7798
QY	8122	ccctgcgtg	8131
Dn	7799	CCCTGCGCTGG	7808
RESULT	8		
AF038601			
LOCUS	AF038601	7333 bp	mRNA
DEFINITION	Sus scrofa porcine endogenous retrovirus ERV-PK15 mRNA, complete sequence.	MAM	21-MAY-1998
ACCESSION	AF038601		
VERSION	AF038601.1	GI:3133304	
KEYWORDS			
SOURCE	pig.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
REFERENCE	Akiyoshi,D.E., Denaro,M., Zhu,H., Greenstein,J.L., Banerjee,P. and Fishman,J.A.		
AUTHORS	Identification of a full-length cDNA for an endogenous retrovirus of miniature swine		
JOURNAL	J. Virol. 72 (5), 4503-4507 (1998)		
MEDLINE	98216827		
REFERENCE	2 (bases 1 to 7333)		
AUTHORS	Fishman,J.A.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (16-DEC-1997) Infectious Disease Unit, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA		
AUTHORS	(bases 1 to 7333)		
REFERENCE	Fishman,J.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-MAY-1998) Infectious Disease Unit, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA		
REMARK	Sequence update by submitter		

COMMENT On May 16, 1998 this sequence version replaced gi:3116445.
 FEATURES
 Location/Qualifiers
 1..7333
 /organism="Sus scrofa"
 /strain="miniature swine"
 /db_xref="taxon:9823"
 /cell_line="PK-15"
 /note="porcine endogenous retrovirus ERY-PK15. Infective virus but represents a naturally occurring transcript"
 LTR
 1..578
 598..2172
 /codon_start=1
 /product="gag protein"
 /protein_id="AAC16766.1"
 /db_xref="GI:3116446"
 /translation="MGQVVTPLSLTLDHMTVSRRAHNLISVQYKKGKPGWOTFCASEMP
 TPDVMPSEGTENSEIILAKAIFQNGPSHPQDEYIILMDLADPPPWKPMIN
 KRKPGPRILAEKNSAKVEPSPIYETIEPTWEPPOVPPYPAQAVAG
 PSAPPGAYVEBPAGTSGRSRGATPERDELAIIPLTYPGPMGGOLOPQWPFSS
 ADLYNMTNHPPESEDPORLTGLVESLMESQPTWDDCOOLLOTLPTERRRILIEA
 KKNVPGADGRPOLONEIDMGEPILTRPCWDYNTABGRSLKTIQALVAGLRASRRP
 TNLAVREVMOGPNPPSPFLERLMEARREPTPDPSSEAKASVALAF7G5SLDIR
 KLIORLEGLOEALDLVLEAEKYYRRETEDEKREKEREERERDRROCKNLT
 KLIANVEKSSSRERERDRFRKIRSGPQSGNLGNRTPLDKDQACAYCEKEGHMANCPK
 KGNKGPVLALEEDMD"
 2320..3552
 /note="truncated protein"
 /codon_start=1
 /product="pol protein"
 /protein_id="AAC16767.1"
 /db_xref="GI:3133305"
 /translation="MGATGCRQVPMWTRRTVYDLAGVTHSEFLVIEPCPVLLRDLL
 TRKGAQISFEOGPEVSNNKRTITVITLQIDDEYLYSPQKPPDODLADPPWVPMIN
 AETAGMGLAKQVPOVYIOLKASATPVSRYOPLHSEABEGHINPHYORLIOGCTLPVQ
 SPWNTEPLVPRKRGNDYRPOVLDREVNKRVDYHPVNPENILSALPBRNMTYLL
 DLKDAFECRLHPTSQLPTFEMRDPGCRGTQKLTWTLPOCFKNSPTFEDALHRLD
 ANFRIOHPVOTLLOYDILLAGATKODILEGTAKLLLELSDIGYRASAKAQCRRR
 VYVIGYSLRGQWMLTEARKTIVVOIPAPPTAKQVREFLGTAGFCRLMIPGFAILAAP
 LYLPLEKESGCLPQGGK"
 4803..6725
 /note="env protein; N terminus deleted"
 /codon_start=1
 /pseudo
 6726..7333
 LTR
 BASE COUNT 1984 a 1788 c 1885 g 1676 t
 ORIGIN

Query Match 36.1%; Score 2967; DB 4; Length 7333;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
 QY 1 gtgtgtgtaagactgtggcccgagcgactgtgaataaaatccctctgtgttgatc 60
 |||||
 Db 11 GTGTGTGTAAGACTGTGGCCCGAGCGCGCTTGAATAAAATCCCTGTGTGTGATC 70
 QY 61 aagaccgcttcctcgtgagtgatggggtgcgcctctccgagcccgagcgagggat 120
 |||||
 Db 71 AAGACCGCTTCTTGTGAGTGAATTTGGGCTGTGCGCTCTTCGAGACCGCGAGAGGGAT 130
 QY 121 tctcttactgagcttcatcttggatggatggcgaggaatccctcgagcaaccctta 180
 |||||
 Db 131 TGTCTTTTACTGGCTTTCATTTGTGCTGTGGCGGAATCCTTGAGACCACTCTTA 190
 QY 181 caaccgagaaacgacttggagtaagagatccctcttggaagctgtgtgttcggcc 240
 |||||
 Db 191 CACCCGAGAAACGACTTGAAGGTAAAGGATCCCTTTGGAAATATGTGTGTGCGGCC 250
 QY 241 ggcgtctcgttctgagtgctgttttcggtatgagcgacttgcggtttgcagctgtc 300
 |||||
 Db 251 GGGCTCTGTTCTGTGAGTGTCTGTTCGATGATGCGGCTTTCGGTTGACGCTGTCT 310
 QY 301 ctgaagacgtgaagactgtgatacgcagacgtgtctgaagagatcaacagctg 360
 |||||
 Db 311 CTCAGACCTAGACGACTGAGACTGTGATCAGCAGAGCTCTAGAGAGATCAGGCTG 370
 QY 361 caacccttgaggagagcccgaggaggttgaggagagcgaagcgctgtgttctctact 420
 |||||
 Db 371 CCACCTCGGGGACGCGCCGAGGTGGGAGAGCGCAGAGAGCGCTGTGTCTCTACT 430
 QY 421 gtcgtaagagagacgagctgtgttctgaagcgaagctccctcccgagcgctgcg 480
 |||||
 Db 431 GTGCGTCAAGAGCAGATTCTGTGTGAAGCAAACTTCCCTCCGCGGCGTCCG 490
 QY 481 actccttgctgctgtgtggaagacgagagcggtgtgtgtgtgtgtgtgtgt 540
 |||||
 Db 491 ACTCTTTCCTGCTGTGTGAAGACGCGGAGCGGTGTGTGTGTGTGTGTGTGTGT 550
 QY 541 ctctcccggtgtcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
 |||||
 Db 551 CTGTTTGCTGTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 610
 QY 601 tgaataccccccttaacttgaacttcgacatctgaactgaactgaactgaactga 660
 |||||
 Db 611 TGACTACCCCTTACTTGTGACTCTGACCAATGAGCTGAAATGATTCAGAGGCTCA 670
 QY 661 attgtcagttcaggttaagaagagccttgcaagcttctgtgtgtgtgtgtgtgt 720
 |||||
 Db 671 ATTGTGCTGCTGAGTTAAGAGGAGCTTGGAGACTTCTGTGTGTGTGTGTGTGTGT 730
 QY 721 catctgctgt 780
 |||||
 Db 731 CATTCGATGTTGATGAGCTTCAAGAGGAGGAGCTTAAATCTGAATTAATCTGCTGT 790
 QY 781 aggaatcaatcttcaagactgaacgagcgtctctatctctgataagagccctatct 840
 |||||
 Db 791 AGGCAATCAATTTTCAACTGAGACCGGCTCTCATCTCTATGAGGAGCCCTTATATCT 850
 QY 841 cgttgcaagatttgagcaagatccctccgcaatggtttaaccatctgaactaaacaa 900
 |||||
 Db 851 CGTGCAAGATTGTGCAAGAGATCTCTCCGCAATGGGTTAAACCATGCTAAATAAACCA 910
 QY 901 gaaagcaggtcccggaatccctgtctcttgagagagaaacaaacactggtgcgaaaga 960
 |||||
 Db 911 GAAAGCCAGGTCGCCGAAATCTGCTGTGTGAGAGAAACAAACACTGCGCGGAAAG 970
 QY 961 tgaagcctctctctgtatctacccgagatcgaagagcgcgagacttgccgaagcc 1020
 |||||
 Db 971 TGAACCTCTC-TCTCTGATATACCCGAGATCAGAGAGCGCGCGACTTGCGCCCT 1029
 QY 1021 caactgttcccccaccccttaccagcaaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1079
 |||||
 Db 1030 CAACTGTCTCCCAACCCCTTATCCAGCACAGGCTGCTGTGAGAGGAGACCTTGCCTCT 1089
 QY 1080 cctgagactccggt 1139
 |||||
 Db 1090 CCTGAGCTCCGCTGTGTGAGAGGACCTGTCTCGGACTGAGACCGGAGAGGCGCCACC 1149
 QY 1140 cggagcgagaaagagagatcgatattacgctgtgcacactatgcccctccatgcca 1199
 |||||
 Db 1150 CCGAGCGGACAGACGAAATCGCATATTACCGTGTGGACATCTGCGCCCTCCATGCCA 1209
 QY 1200 gggggccaattgcaagccctccagatattgagccttctctctgagatctctataat 1259
 |||||
 Db 1210 GGGGGCAATTTCAGACCCCTCCAGATATGCGCTTCTTCTGAGATCTCTATAAATGG 1269
 QY 1260 aaaaactaaatcccccttctcgaagatccccaagcctcgaagaggtgtgtgtgtgt 1319
 |||||
 Db 1270 AAAAATCAACATCCCTCTTCTGAGAGATCCCAACGCCCTCAGAGGGTGTGTGAATCC 1329
 QY 1320 ctatgtctctcaacagctacttggatgtatgttcaaacagctgtgtgtgtgtgtgtgt 1379
 |||||
 Db 1330 CTATGTCTCTCACAGGCTACTTGGGATGATTTCAACAGCTGTGCAACACTCTCTTC 1389
 QY 1380 acaacgagagagagagagaaattctgttaagaggttaaataaaagtccctgtgtgtgt 1439
 |||||

Db 1390 ACAACGAGAGGAGAGAGAAATTTGTTAGAGCTTAATAAATGTTCTTGAGGCGCGAC 1449
 QY 1440 gggcgagccagcgagctgcaaaatgagatlgaca tgggattcccttgatcgcccggt 1499
 Db 1450 GGGCGAGCCGACGACTTCAAAATGAGATGACATGGGATTTGCTTGACTCCGCCGCT 1509
 QY 1500 tgggactaacacagcgagtgagagagcgcttgaataatctatcgcaagctctgltg 1559
 Db 1510 TGGGACTACACACGCGCTGAAAGGTAGGAGGCTTGAAATCTATCCCGCTGCTGTTG 1569
 QY 1560 gcgggctccggggcgccctcaaga cggcccaataatttgcttaaggaagaggtgag 1619
 Db 1570 GCGGGTCTCCGGGCGCGCTCAAGACGGCCCACTAATTTGGCTTAAGGTAAAGGTGATG 1629
 QY 1620 caggagccgaaacgaactccctcgatattcttgaagcgctcaatggaagccttcaagcg 1679
 Db 1630 CAGGACCGCAAGCAACTCCCTCGATTTCTTGAGAGGCTCATGGAAGGCTTCAGGCGG 1689
 QY 1680 ttcaacccttctgactcctacactcagagagccagaagaagcctcagtgccctgctcat 1739
 Db 1690 TTCAACCCCTTTGATCTTACTACTAGAGGCGCCAGAAAGCCTCACTGGCCTTGCGCTTCAT 1749
 QY 1740 gggcagtcgagcttgatatacagaagaactcagaagactggaaggttacaagagct 1799
 Db 1750 GGGCAGTCGGCTCTGATATCAGGAAGAACTTCAGAGACTGGAAGGCTTACAGAGCT 1809
 QY 1800 gagttagctatctagtgagagagagagagagaggttatatacagaagggagacagaag 1859
 Db 1810 GAGTACTGATCTAGTAGAGAGAGAGAGAGAGAGGTATTCAGAGAGGAGACAGAGAG 1869
 QY 1860 gagaaggaacagagaagaagaagaagagagagagagagagagagagagagagagag 1919
 Db 1870 GAAAGAGACAG 1929
 QY 1920 caagagaagaatttgactagaatcttggccgagctggttgaaggaagagagagagagag 1979
 Db 1930 CAAGAGAGAAATTTGACTAGATCTTGCGCGACAGTGTGAAGGAGAGAGAGAGAGAG 1989
 QY 1980 agagagagagattttagaanaattagtgcaagccctagaagctcaggaagacccggcaat 2039
 Db 1990 AGAGAGAGAGATTTTAGAAATTTAGTTCAGGCGCTTACAGAGGTCAAGGAGAGAGAG 2049
 QY 2040 aggagcccccctcgacaagaagcagtgctgctatgttaagaagaagaacacttggcaag 2099
 Db 2050 AG 2109
 QY 2100 aactgcccccaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2159
 Db 2110 AACTGCCCCAG 2169
 QY 2160 tagggagagagaggttgcgaacccctcccgagcccgaggttaacttgaagaggtggaggg 2219
 Db 2170 TAGGGAGAGAGGGGTTGCGAGCCCCCTCCGAGCCAGGAGGTAACTTTGAAGGTGAGGGG 2229
 QY 2220 caacgaattgagttccttggtgataccggagcggagcattcagtgctgtacaaacatta 2279
 Db 2230 CAACGAGTTGAGTTCTGCTGGTATACCGGAGCGGACATTCAGTGTCTTACAAACATTA 2289
 QY 2280 ggaagaactaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2339
 Db 2290 GGAAGAAGTAAAG 2349
 QY 2340 tggactaccggaagaacgcttgagcttggaggtggaacgggtlaacccactcgtttctgggc 2399
 Db 2350 TGGACTACCGAG 2409
 QY 2400 atcccttagtgcccaagtaacccctctagtgagagagacttctcaagaagaaggagatcaa 2459
 Db 2410 ATCCCTAGTGCCCAAGTACCCCTTTAGGTAGAGACTTACTACCAAGATGAGAGACTCAA 2469
 QY 2460 attcttltgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2519
 Db 2470 ATTCTTTTGAACAG 2529

QY 2520 accctcaattagatgatalcagatatalcttccccaagtaagagctgaatcaagat 2579
 Db 2530 ACCCTCAATTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2589
 QY 2580 ataacgtccttggttggagcgagttccccaagcgcttggcagaagaacgcaaggatgggt 2639
 Db 2590 ATACAGTCTCTGTTGGAGCGAGTGTTCGCCAAGCGCTGGGCGACAAACCCAGAGATGGTTG 2649
 QY 2640 gcaagaagagttcccccacaggttatcaactgaagcgcaagctctacacagatcaagc 2699
 Db 2650 GCAAGAAGCAAGTTCGCCACAGGATATTCACTAGAGGCCAGTCTTACACAGATCAAGTC 2709
 QY 2700 agacagtaaccccttgatgaagaaggtcgagaagaatttggccgcatgtlcaaatga 2759
 Db 2710 AGACAGTACCCCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2769
 QY 2760 atccaagaaggacatccagctcctgctcgaatcccttgatattccctgtacaggt 2819
 Db 2770 ATCCAAGAGGACATCCAGTCTCTGCTGCAATCCCTTGGAATATCCTCTGCTACCGGTT 2829
 QY 2820 aggaagccttggaacaaatgatalcgaaccagtaagagacttgaagaggtacaataaagg 2879
 Db 2830 AGGAAGCCTGGGAGCAATGATTTACAGCAGTACAGAGACTTGAAGAGTCAATTAAGG 2889
 QY 2880 gtgaagagatacaccacaacggtcccgaaacccctataaccccttgagcgccctccgct 2939
 Db 2890 GTGACAGACATACACCCCAAGGTCGGAACCTTTATACCTCTTGAGGCGCTCCGCT 2949
 QY 2940 gaaaggaactgtatcacaglatlgtgaactaagaatgctcttctgtgcttgaatlaac 2999
 Db 2950 GAAGGAGACGTGATACAGTATTTGACTTAAAGATGCTTCTTCTGCTGATGATAC 3009
 QY 3000 cccactagcaacacacttcttgcttgaatgaagaatccaggaaccccgacacatttgaagaa 3059
 Db 3010 CCCACTAGCAACCACTTTTACTTTCATGAGAGAGATCCAGTACGGAAGAACCGGG 3069
 QY 3060 cagctcaacttgagccagacttgcccaaggttcaagaagctcccgacacatttgaagaa 3119
 Db 3070 CAGCTCACTGTGACCCGAGTCCGAGTCCCAAGGTTCAAGAACTCCGACCATCTTGGACGA 3129
 QY 3120 gccctacacagagagacttggccaacttcagatlaacaaacccctcagtgacccctccag 3179
 Db 3130 GCCCTACACAGGAGACTTGCCCAACTTCAGATCCAAACCCCTCAGTGAGACCTCTCCAG 3189
 QY 3180 tacgttgatgacgttctcttgcttggcgagcccaacaaacagagacttgtagaagtaagaa 3239
 Db 3190 TACGTGATGACCTGCTCTGCGGAGGCCAACAAACAGAGACTGCTTAGAAGTACGAAG 3249
 QY 3240 gcaactacttgatattgtcgaactaggtctacagagcctctgctaagagagccagatt 3299
 Db 3250 GCACACTAGCTGGAATGCTGAGACTTAGGCTACAGAGCCTCTCTTAGAAGAGGCCAGATT 3309
 QY 3300 tgcagggagaggtgatacactatgggttacagtttgcggggcgagcgagatgctgacg 3359
 Db 3310 TGCAGGAGAGGTAACTACTTGGGTTACAGTTTGGGGGCGGAGCGAGATGGCTGAGG 3369
 QY 3360 gaggcaggaagaagaacttgatccagatccgggcccccaacccagccaaagaagtgaga 3419
 Db 3370 GAGGCAAGGAG 3429
 QY 3420 gagtlttggggagagcttgatattgcagacttggatccggggttggcagacttaaga 3479
 Db 3430 GAGTGTGGGGAGACAGCTGATTTTGCAGACTGTGATCCGGGGTTTGCAGACTTACGA 3489
 QY 3480 gccccaacttaccgctcaacccaagaagaagaagggg 3514
 Db 3490 GCCCACTTACCCGCTTAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3524

RESULT 9
 ARL130474
 LOCUS ARL130474 7333 bp DNA PAT 16-MAY-2001

Db 81 GAGAGATTCAGGATGCGGAGAAACCGGACACTGACCTGACCGAATGGCCCAAGGTT 140
QY 3092 caagaactcccgagacactcttgagagagccctacacaggaactggccaactcagat 3151
Db 141 CAAAGACTCCCGGACCACTTTTGACGAAGCCCTACACAGAGGAGCTGGCCAACTTCAAGAT 200
QY 3152 ccaacaccccttaggtgacctccctccagtaagtgatgacctgtcttgccggagagcac 3211
Db 201 CCAACACCCCTCAGGTGACCCCTCCCTCAGTACGTGATGACCTGCTTGTGGCGGGAGCCAC 260
QY 3212 caacacagactgcttagaaggtacagagcaactactgctgaattgtctgacctaggta 3271
Db 261 CAACAGAGACTGCTTAGAGAGGTAGAGAGGACACTACTGCTGAATTTGCTGACCTAGCCTA 320
QY 3272 cagagcctctgcttagaagagccagatcttcagagagagaggtataactctgggtgacag 3331
Db 321 CAGAGCCTCTGCTAAGAGGCGCCAGATTGTCAGAGAGAGAGGTAACTTGGGGTACAG 380
QY 3332 ttctcgaggcgagagcagatgctgacgagagcagagaaagaaactgtagtccagatacc 3391
Db 381 TTGCGGGGGCGGACGATGGCTGACGAGGACGAGAGAAACTGTAGTCCAGATACC 440
QY 3392 ggccccaacacagccaacaaacagatgagaggttttgaggacagctgatttgcaact 3451
Db 441 GGCCCAACACAGGCAACAAAGTAGAGAGATTGTTGGGACAGCTGATTTGCAAGCT 500
QY 3452 gtgatacccgaggtttgagaccttagagcccaactctacccgcttaacccaagaagaag 3511
Db 501 GTGATCCCGGGGTTTGAGACCTTAGACGCCCACTCTACCCGCTAACCAAGAAAGAG 560
QY 3512 gggatctccctgggctccttgagaccaagaagcatctgactcctacaaagagccctgct 3571
Db 561 GGGATTTCTCTGGGCTCTTGAGCACCAAGAGCATTTGATCTTCAAAAAGGCCCTGCT 620
QY 3572 gagcagacctgctctgacctcccttgacgtactaaaccccttaaccttaagtgtgata 3631
Db 621 GAGCCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCTTTACCTTTATGTGATGA 680
QY 3632 ggcgtaaagagtagcccgagaggttttaacccaaccccttagacactgagagagacgt 3691
Db 681 GCGTAGGGAGTAGGCCCAAGAGTTTAAACCAACCCCTAGACCATGAGAGAGACCTGT 740
QY 3692 tgcctacctgtcaagaagactgtacctgttagccagtggtgtggccgctatctctgaagc 3751
Db 741 TGCCACTCTGTAAAGAGCTTGATCTGTAGCCAGTGGTGGCCCGATGTCTGAAGGC 800
QY 3752 tatcgcaagctgtggccatactggtcaagagcgtgacaaattgaccttggagcagata 3811
Db 801 TATCGACACTGTGGCCATACTGTGTCAAGAGCGCTGACAAATTTGACTTTGGGACAGAAAT 860
QY 3812 aacttaagccccccctgactgtagagaatactgttgagagcccccaagccagatgact 3871
Db 861 AACTGTAAATAGCCCCCATAGCATTTGGAACACTGCTTGGCGGCCCCAGACCATGTGAT 920
QY 3872 gaccaagcccccgatgacccaactacaaagcctgtctctcaagagaggtgcaacttgc 3931
Db 921 GACCAAGCGCGCATGACCCACTATCAAAAGCCTGCTTCTCACAGAGAGGTCACATTGGC 980
QY 3932 tccacacagcagctctcaacccctgcaactctctcgtcgtgaagagactgtgaacagtgac 3991
Db 981 TCCACACAGCGCTGTCAACCCCTGACACTTCTGCTTCAAGAGATGTGAACAGTATAC 1040
QY 3992 taatattgccaactaactatgattgagagagactgggttcccaagagcccttaagacat 4051
Db 1041 TATATGTTCCATCAACTATTGATTGAGAGAGACTGGGGTCCCAAGAGACCTTTACAGATAT 1100
QY 4052 accgctgactggaagaagtgtaactggtctacgtgacggaagcagctatgtgtgtgaag 4111
Db 1101 ACCGCTGACTGGAGAGTGTCTAACTGTTTCACTGACGGAACACACTATGTGTGTGAAG 1160
QY 4112 taagagatgcttgaggcgagtggtgtgagcggaacccgcaagatctgtggccagagact 4171
Db 1161 TAAGAGATGCTGTGGCGGCAAGTGTGAGCGGGAGCCCGACAGATCTGGGCGACAGCCT 1220
QY 4172 gccggaaggaacttaagcgcaaaagcttgagctcatgtgacctcaagcaagcttgagct 4231
Db 1221 GCCGGAAGGAACCTTCAAGCCCAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGGGCT 1280
QY 4232 gccgaaggaataccataaactatatacggacagcaggtatgaccttgagctgacaa 4291
Db 1281 GGCCGAAGGAATTCATTAACATTATATACGACACAGCTATAGCTTTGGCAGCTGACAA 1340
QY 4292 cgtacacggggccaactataaagaaggggtgtgtcttaccttaagcagggaggaataaa 4351
Db 1341 CGTACACGGGGCCATCTTATAAACAAGGGGCTGTCTTACCTCAAGCGGGAAGAAATTA 1400
QY 4352 gaaacaaaggaatcttcaagcctatagaagccttacctatcttgccaaaaagctagctat 4411
Db 1401 GAACAAAGAGAAATTTCTAAGCTATTAGAAGCTTACATTTGGCCAAAAGGCTAGCTAT 1460
QY 4412 tatacactgtcctgagacatcagaagaagccaagaatctcatactataagaggaaacatgag 4471
Db 1461 TATACACTGCTGAGACATCAGAAAGCCAAAGATCTCATATCTAGAGGAGACCAAGTGGC 1520
QY 4472 tgaacgggttgccaagcagcagccaggtgtttaaactcttgacctataatagaacgac 4531
Db 1521 TGACCGGGTGGCCAAAGAGGAGCCAGCCAGGCTGTTAACTTTGCTTATATACAAACGCT 1580
QY 4532 caaagccccagaacccagacagtlacacccctagaagactggcacaagataaaaaagat 4591
Db 1581 CAAAGCCCCAGAACCCAGACAGACAGTACACCTTAGAAGACTGCGCAAGAGATAAAGAAAT 1640
QY 4592 agaacagctctctgagaactccggagaggagctcttacctataatgggaagaatactct 4651
Db 1641 AGACAGATTCTCTGAGACTCTCGAGGGGAGGACTCTTATACCTCATATATGGGAAGAAATCT 1700
QY 4652 gccccaagaagaggttgaataatgtccaaagataacatcgtctaaaccaactagaagac 4711
Db 1701 GCCCACAAGAAAGAGGTTGATAATATGTCCAAAGATATCATGCTTAACCCACTAGAGAC 1760
QY 4712 taaacactctgacagagttgtgtcagaacatcccttactcatgtctcagagctaccagag 4771
Db 1761 TAAACACTCTGACAGAGTTGGTCAGAAACATCCCTTATCATATGTTGAGGCTACAGAGAT 1820
QY 4772 ggtgactcgttggttcaaacatgtgtgtccctcgaagctgtgttaagttaactctccag 4831
Db 1821 GGCTGACTCGGTGTCAAACATTGTGTGCTCCCACTGCTTAATGCTTAATCTTCCAG 1880
QY 4832 aatacctcaggaagaagacttaaggggagccacccagcgctcaactgaggaagtggact 4891
Db 1881 AATACCTCCAGAAAGAGACTTAAGGGAAGCCACAGGCGCTACAGGGAAGTGGACTT 1940
QY 4892 cactgaggtlaagccggtctaaatacaggaacaaatactatgtgtttgttagaacctt 4951
Db 1941 CACTGAGGTAAAGCGGCTTAATATGCGGAACAAATATCTATGTTGTTTGTAGAACCTT 2000
QY 4952 ttcaagatggtagaggtctatctactaagaagaagacttcaacgltgtgtctagaag 5011
Db 2001 TTCAAGATGGGTAAAGGCTTATCTTAAAGAAAGAACTTCAACCGTGTGTGTAGAA 2060
QY 5012 aatactgaggaatcttccaaagatttgaaataccaaagtaagtaagtgctcaagaatg 5071
Db 2061 AATACCTGGAGGAATTTTCCAAAGATTTGGAATATCTAAGAGTAATAGGTCAGACAAATG 2120
QY 5072 tccagactctgttgccaggtlaagtcaggaacttgcccaagatalatgggactgtattgaa 5131
Db 2121 TCCAGCTTGTGCTTCCAGGTAATGTCAAGGAGCTGCGCAAAATVTTGGGGATTATAGGA 2180
QY 5132 actgcatctgtgcatatagagcccaagcctgaagcagtagagagatagaagaacat 5191
Db 2181 ACTCATTTGTGCATACAGACCCCAAAAGCTCAAGACAGGTGAGAGGATGATGAACCAT 2240
QY 5192 taaagagaccccttaacaaattgacacagagactggcatatgattgatagtgctcct 5251
Db 2241 TAAAGAGACCTTACCAAAATTTGACACAGAGACTGGCATTAATGATTTGATGGCTCTCT 2300

	Query Match	Best local Similarity	24.8%; Score 2034; DB 4; Length 3025;	99.7%; Pred. No. 0;	Matches 2574; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY	1 gfgvgltacgactgtgagccacgcgcgttggaaataacatcctctgtcgttgcata	445 GGGGTGTACGACTGTGGGCCCCACGCGCCTTGGAATAAATATCCCTGTGTTGATC	60		
QY	61 aagacgcgtctcgtgagtgatcttggggltgcgcctctccgaagcggagagggagat	505 AAGACCGCTTCTCGTGATGATTTTGGGGTGTGCGCTTCTCGAGCGCGAGAGGGGGAT	120		
QY	121 tglctcttactgctgcttcatcttggltgcgttgcgcgggaatcctgcgaccacctta	565 TGTCTTTTACTGGCCTTTATTTGCTGCTGTGGCGGGGAATCCTGACACCCCTTA	180		
QY	181 gacccggagaacgcgacttggaggtaaaggatccctcttggaaagtgtgtgtgcggcc	625 CACCCGAGAACCGACTTGGAGTAAAGGATGCCCTTTGGAAGATGTGTGTGTCGGCC	240		
QY	241 ggcgcctctgttcttaagtgtcgttcttcggtgatgcgcgtcttcgattgagcttgcct	685 GCGGCTGTGTTCTGAGTGTGCTGTTTGGGTGGATGCCGCTTGGCTTGCAGCTGTGCT	300		
QY	301 ctcaagaccgtaaggaactggaagactgtgatacgaagacgtctcagaagatcacaggctg	745 CTCAAGACCGTAAAGGACTGTGAGAGACTGTATATGACGAGCTGTCTAGGAGATCACAGGCTG	360		
QY	361 caacccttggagagccccccggagaggttggggagagccaaaggagcgttgttctcact	805 CCACCTGTGGGGAGACCCCGGGGAGGTGGGGAGAGCCAGAGACGCTTGCTCTACT	420		
QY	421 gtccgtcaagagaccgagltctgttgttgaagcgaaagcttcccccctcgcgcgtccg	865 GTCCGTCAAGAGACCGAATCTGTGTGTGTAAGCAAAAGCTTCCCCCTCCGGCGCGTCCG	480		
QY	481 acctcttgcctgtcttggaaagacggagaggttcgctgtgtcgtgactgttgttctt	925 ACTCTTTGCTCGCTGTGGAGAGAGCGGAGCGGCTCGGTGTGTGATCTGTTGGTTT	540		
QY	541 ctgtctcgttgttctcttcttctgtgcgtcttgccttgcataagtttaaatgtgacaagag	985 CTGTTCGTGTGTTGTTGTTGTGTGCGCTTGTCTACAGTTTAAATATGGACAGACAG	600		
QY	601 tgaatacccccttagttgactctgacaatltgagactgaagttagatcacagggtcata	1045 TGACAGACCCCTTAGTTGTGACTCTGACCACTTGTGACTGAAGTATGATCCAGGGCTCATATA	660		
QY	661 atttgcagttcaagttaaagaagagacttggagacttctgtgcctcgtgaatggccaa	1105 ATTGTGATGTTCAAGTTAAGAAAGGACCTTGGCAGACTTCTGTGCTCTGAAAGGGCCAA	720		
QY	721 catcgaatglttggatgtggccatcagaagggagaccttaattctgaaatatatcctgtccta		780		

|||||
Db 1621 GAAGGCGGCTTATGTGAGCCTTAAAGAAATGTTGCTTATGTATGATCACTACAGCA 1680
Qy 7300 gccataagagactccatgacaagcttagaaaaaagtagagagcgctgaaggaaga 7359
Db 1681 GCCATGAGACACTCCATGAAACAGCTTAGAAAAAGTTAGAGGCGGTGAAAGGGAAGA 1740
Qy 7360 gaagctgacacagggggtgttgaagatgttcaacagctctccttgatgacacccg 7419
Db 1741 GAGGCTGACCAAGGGGTGTTGAAGATGTTCAACAGCTCTCTTGATGACCACTCG 1800
Qy 7420 cttctgctcgaacggggccctagtagctgctcctgcttacttaacgttgagcctgc 7479
Db 1801 CTTTCTGCTGACGAGGGGCCCTAGTAGCTCTCTCTGTTACTTACAGTTGGGCTTGC 1860
Qy 7480 ttataataagattgttgctccttgtagagaaacagtagagcaatccagatcatgta 7539
Db 1861 TTAAATTAATGTTTGTGCTTGTGTTAGAGAACGAGTAGTCAGTCCAGATCATGTGA 1920
Qy 7540 cttaagcaacagtagcaagccttcgaagcaagagaagaaactgaactctag 7590
Db 1921 CTTAGGCAACAGTACCAAGGCTTCTGAGCCAGAGAAACTGACCTTAG 1971

RESULT 13
A66559 1974 bp DNA PAT 29-MAR-1999
LOCUS A66559
DEFINITION Sequence 9 from Patent WO9740167.
ACCESSION A66559
VERSION A66559.1 GI:4538112
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 1974)
AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
TITLE PORCINE RETROVIRUS
JOURNAL Patent: WO 9740167-A 9 30-OCT-1997;
Q ONE BIOTECH LTD (GB)

FEATURES
Source location/Qualifiers
1..1974

BASE COUNT 534 a 489 c 483 g 468 t
ORIGIN
/organism="unidentified"
/db_xref="taxon:32644"

Query Match 18.3%; Score 1499; DB 6; Length 1974;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

Qy 5639 ggcgcacacccgcgactcgggtgtgagagcgcaaaagactgaatcccttaagctg 5698
Db 20 GGGCCACCTCCGACCTGGGGTGGAGAGCCGAAAGACTGGAATCCCTTAAAGCTTCG 79
Qy 5699 cctccacgcctgttccctactcctaaactcccccagcgagtagtaaacgcctta 5758
Db 80 CCTCCATCGGCTGGTTCCTTACTCTAATCAATACCTCCCGGCAAGTAAACGGCTTA 139
Qy 5759 taagacagctgaacccccaatagaccttalccttaccctgagctgattatgacctgata 5818
Db 140 TAGACACTCGAACCCCAATAGACCTTATCCCTACCTGCGATATATGACCCGTGATA 199
Qy 5819 cgggtgtacatgtaaatagacacagagtgctgctccttagagcactgtgtgcgaac 5878
Db 200 CGGGTGTACCTGTAATAAGACCTGAGGTGTGCTCTTAGAGGCACTGGTGGCTGAAAC 259
Qy 5879 tgacattctgcctcgaatgatalaacccgcgtgtlaaaagacaacccccaactagctc 5938
Db 260 TGCATTTCTGGCTCCGATGTGATTAAACCCGCTGTAAAGACACACCTCCCAACCTAGTCC 319
Qy 5939 gtagattgggtctctatgtctgcacagacagagaagaataactgtgggggtctcg 5998
|||||

Db 320 GTAGTTATGGGCTCTATTGCTGCCAGCCACAGAGAAAGAAATACTGTGGGGTCTCG 379
Qy 5999 gggaaatcctctgtagagatgagagctgctgacccccaacagatgagagctggaatgac 6058
Db 380 GGGATTCCTTCTGTGAGAGATGAGACTCGTCACTCCCAAGATGAGAGATGGAATGCG 439
Qy 6059 cgaatcctcacaagaccgggtaaatctcctctgtcattccggccgggcaagtaea 6118
Db 440 CGATCTCTCTCCAGAGCCGGGTAATATCTCTTGTCAATTCGCGCCGGGCAAGTACA 499
Qy 6119 aatgataaactatataagaatagactctccccaatagacttagatattatataaga 6178
Db 500 AATATGATAACTATATTAAGATTAAGACTGCTCCCAATCAAGACTTAATTAATTAAGA 559
Qy 6179 taagttcaactg--aaaagaaacagaaatattcaaaagtgtgaataatgttagact 6236
Db 560 TAAAGTTTCACTAAAGAAAGAAACAGAAATATTCAAAAGTGGATTAATGTATGACT 619
Qy 6237 ggggaatagttttat--tataatgacggggaagglccaacttaacatlcgacctta 6295
Db 620 GGGGAATGATTTTATTAATATGCGGGGAGCAGGGGTCCACTTAAACCAATTCGCTTA 679
Qy 6296 gatatagagacggggaacgaacccctgtgtgaaatgtgagcccgataaagtaactgtgac 6355
Db 680 GGATGAGACGGGGACAGAACCCCTGTGGCAATGGACCGGATTAAGTACTGGCTGAC 739
Qy 6356 aagggcccccggccctgtgagccacgcataactggtgcgtgcccccaatlaactgcctgc 6415
Db 740 AAGGGCCCCCGGCTTGAGAGCCACCGCATTAATCTCGCGTCCCAATTAACCTCGCTGC 799
Qy 6416 ggcctgacataacacagccgccttagcaaacagtaacatgtagtattcttcccaacacgc 6475
Db 800 GGCCTGACATAACACAGCGCCCTTACCAACAGTACACATGATGATTCCTTCAACACAGC 859
Qy 6476 ctagaacatccccagtgcttctgttgaagaagcaagactctcagatctatcaag 6535
Db 860 CTAGAAACCTCCCAAGGTGTCTGTTAAGACGGGACAGAGACTCTTCAAGTCAATCCAGG 919
Qy 6536 ggccttcccaagcacaactcaccacgcacccctgaagcaactctctgtgtgtctgc 6595
Db 920 GAGCTTTCACACCATCAACTCACCAGACCGCATGCGCACTTCTTGTGGCTTGTGC 979
Qy 6596 tatccatagggccctccttattatgaaggatgagctaaagaaatcaatgtagcca 6655
Db 980 TATCCTCAGGCGCTCCTTATATGAGGGGATGGCTAAAGAAAGAAATTCATATGACCA 1039
Qy 6656 aagagcatagaataaagttagatagtggtgtccggaataagtaacctcactgaagttt 6715
Db 1040 AAGAGCATAGAAATCAATGTACATGGGGGTCCGAAATTAAGCTTAACCTCACTGAATTT 1099
Qy 6716 ccgggaaggagacatgacataggaagaagctccccaatcccaacacacttgctatgla 6775
Db 1100 CCGGGAAGGGGACATGACATAGGAAGAAAGCTCCGCCATCCACCAACACCTTGTCTATAGA 1159
Qy 6776 ctgtgttataagagagcctcagaanaatcaglattagtlacgtgttataacagtgct 6835
Db 1160 CTGTGTTTATGAGAGCGGCTCAGAAATCACTATTTAGTACTGTTATTAACAGTGTGT 1219
Qy 6836 gggatgcaatctggtttaaaccctgtgttccacactcagcttcaaccaatcaag 6895
Db 1220 GGGCATGCAATCTGGGTAAACCCCTGTGTTTCACTCACTTTCATCAACCATCAAG 1279
Qy 6896 attctgtatcatggtcccaatcgtcccccagtgtagtaccatcctcgaagaagtgtcc 6955
Db 1280 ATTGTGTCTCATGTGTCGCAATCGTCCCGAGTGTACTACCATCTCGGAGAGTGTC 1339
Qy 6956 ttgataatatagactatcggtataacagcaaaaagagaaccgtagtacccttaccctag 7015
Db 1340 TTGATGAATATGACTATCGGTATTAACCGACCAAAAAGAGAACCCGTATCCCTTACCTAG 1399
Qy 7016 ctgtatgtcctggaatagagagcggtgtgtgtgtgaagaacagagacgtgacctgata 7075
Db 1400 CTGTATGTCTCGGATTAAGGAGCGCGCTTGGCTAGGAACAGGACAGCTGCCCTGATCA 1459
|||||

Db	990	CCTCACATGCGCTGGTTCCTCTACTTAACATAAATCTCCCGAGGCGAGTAAAGGCGTTA	104
QY	5759	tagaaagctgaaacccccatagacccttatacccttaacctgctgattatagacctgata	5818
Db	1050	TAGACAGCTCGAACCCCCATAGACCTTTATCCCTACCTGGCGATTATTTAGACCTGTATA	1109
QY	5819	cgggctgaactgtaataagacactggaagtgctctcctagaagcactggttgacctgac	5878
Db	1110	CGGGTGTACGTGTAAATAGCACTGGAAGTGTCTCTCTAGAGCACTGTGTGCTTGAC	1168
QY	5879	tgcattctgcctccgatltgattaaccccgctgtltaaaagacacactcccaactagtc	5938
Db	1170	TGCATTTCCTGCGCTCGGATGTATTAAACCCCGCTGTAAAGACACCTCCACACTAGTC	1229
QY	5939	gtagttatgggtttctatctgctgcgcaggaacagagaagaatactgttggggcttg	5998
Db	1230	GTAGTTATGGGTTCTTATTTGCTGCCAGGCAACAGAAAGAAATACTGTGGGGTTCG	1289
QY	5999	gggaatccctctgttaggaagatlgagagctgggttaacccccaagatlgagctggaaatgc	6058
Db	1290	GGGAAATCCTTGTAGAGGATGGAGCTGCGTCACTCAACATGGAAGACTGGAATATGC	1349
QY	6059	cgatctcctlccagagaccgggtlaaatacttccctcttgcaattccgcccgggcaagtaca	6118
Db	1350	CGATCTCTCTCGACAGACCGGGTAAATATTCCTTTTGTCAATTCGCGCCGGGCAAGTACA	1409
QY	6119	aaatgataaactataaagaataaagctgctcccatcagactagattactaaaga	6178
Db	1410	AAGTATGTAACCTATTATAAAGATTAAGAGCTGCTCCCATCAACATTAAGATTATTAAGA	1469
QY	6179	taagtttcacg--aaaggaaaacaggaaaataatttcaaaagtggataaattglatagact	6236
Db	1470	TAACTTCTACGAAAAAGAAAAGAAAACAGAAAATTTCTAAAGTGATTAATGATAGACT	1529
QY	6237	ggggaatagtttttat-tatatggcggggagcaaggtgcacttlaacacttcgacctta	6295
Db	1530	GGGGAATAGTTTTTTATTAATATGGCGGGGAGACAGGCTCCACTTAAACATTGCCCTTA	1589
QY	6296	ggatagagacggggagacagaaacccccctgtgycatgtggaacgcgataaagttaactgctgac	6355
Db	1590	GGATAGAGAGCGGGGACGAGAACCCCTGTGGAGTGGGACCGGATAAGTACGGGCTGAC	1649
QY	6356	agggggcccccgccctggagccacgcgataacttgcggtgtgcccaatlaacctgcgtgc	6415
Db	1650	AGGGGCCCCCGCGCCCTGGAGCCACCCCATTAACCTTGGCGGTGCCCAATTAACCTCGCTGC	1709
QY	6416	ggctgtgacataaacacagccgcttagaagaagataaacactggaattgtttctaccacaacgc	6475
Db	1710	GGCTTGACATTAACACACCCGCTAGACAGGATACCACTGGATTGTTCTTACCAACACGC	1769
QY	6476	ctagaaacctcccaagctgttctcgttlaaagacaggaacagactcttaagtcataccag	6535
Db	1770	CTAACAACCTCCCAAGGTGTCCGTGTTAACAGAGACAGACTCTTCAAGTCTATCCAGG	1829
QY	6536	gaagcttcacaagccatcaactcocaacgacccctgatlgaactctctgtgtgctgtgc	6595
Db	1830	GAGCTTTCGAAGCAATCAACTCACCGACCTCGATGCCACTTCTTGTGTGGCTTGTGC	1889
QY	6596	tatcctcagggtcctccttatatagagggaatgctctaaagaagaagaataatcaatgtgaca	6655
Db	1890	TATCTCTCAAGGCGCTCTTATTATAGAGGGATGGCTTAAAGAGGAAATTCATCTGACCA	1949
QY	6656	aagagcaataaataaataatcatatgggggttcccggaataaagcttaacctcaactgaagttc	6715
Db	1950	AAGAGCTAGAAATCATATGATCGAGGGGTCCCGAAATTAACCTTACCTCATGAAATTT	2009
QY	6716	ccgggaagggagacatgcatagagaaagctcccccataccacaacacacttgcataagta	6775
Db	2010	CCGGGAAGGGGACATGCATATAGGAAAGCTCCCCCATCTCCACACACACTTTGGCATATGTA	2069
QY	6776	ctgtggtttatagcagcgcctcagaataatcaglatattagtaactgtgtataaacagtggt	6835
Db	2070	CTGTGGTTATATGACAGGCGCTCAGAAATCATGATTATTGTACCTGGTTAATTAACAGGTGT	2129

Qy	6636	ggagatcaactactggtttaaaccctctgttctcaactcagtcctcaacaaacgaag	6895
Db	2130	gggagatcaactactggtttaaaccctctgttctcaactcagtcctcaacaaacgaag	2189
Qy	6696	attctctgtctcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	6955
Db	2190	attctctgtctcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2249
Qy	6956	ttgataatctcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7015
Db	2250	ttgataatctcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2309
Qy	7016	ctgtaactcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7075
Db	2310	ctgtaactcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2369
Qy	7076	caggaacacagcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7135
Db	2370	caggaacacagcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2429
Qy	7136	tcgagacctaagaagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7195
Db	2430	tcgagacctaagaagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2489
Qy	7196	tggttctcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7255
Db	2490	tggttctcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2549
Qy	7256	cagccttaagaagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7315
Db	2550	cagccttaagaagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2609
Qy	7316	tgaaacagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7375
Db	2610	tgaaacagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2669
Qy	7376	ggttctcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7435
Db	2670	ggttctcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2729
Qy	7436	ggcctcagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7495
Db	2730	ggcctcagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2789
Qy	7496	ttgctctcagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7555
Db	2790	ttgctctcagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2849
Qy	7556	aagcctcagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7615
Db	2850	aagcctcagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2909
Qy	7616	attaacagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7675
Db	2910	attaacagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	2969
Qy	7676	ccaggaacacagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7735
Db	2970	ccaggaacacagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	3029
Qy	7736	aaatagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7795
Db	3030	aaatagtcaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	3089
Qy	7796	ataacaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7855
Db	3090	ataacaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	3149
Qy	7856	agaacaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7915
Db	3150	agaacaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	3209
Qy	7916	tgccaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	7975
Db	3210	tgccaggaacacagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	3269
Qy	7976	agcgcgagtcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	8034
Db	3270	agcgcgagtcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	3329
Qy	8036	taaatagtcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	8094
Db	3330	taaatagtcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	3389
Qy	8096	tcgcagtcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	8154
Db	3390	tcgcagtcagtcgttcaaatctgtcccgagtgtaactcaactcttgaggaagtggttc	3449
Qy	8156	gcttgaataaatactctctgttgcacac	8187
Db	3450	gcttgaataaatactctctgttgcacac	3482
RESULT 15			
PERENV2			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
FEATURES			
source			
gene			
CDS			
BASE COUNT			
ORIGIN			

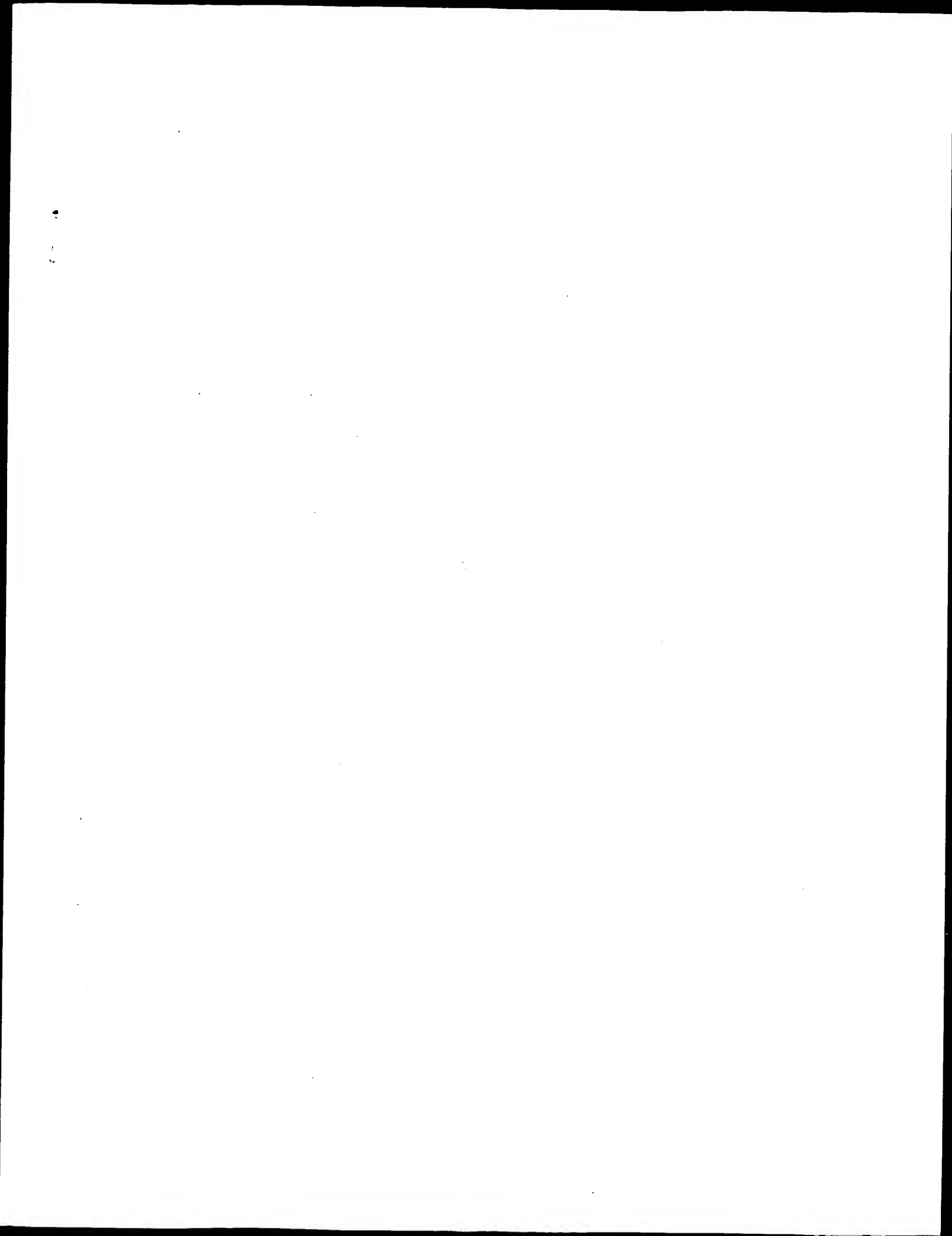
Query Match 17.9% Score 1466: DB 14; Length 3482;
Best Local Similarity 99.3% Pref. No. 0;
Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

QY 5639 ggcgcacactccgcagctcgagggtgagagccgaagaagctgagaatcccttaagctcg 5698
DB 930 GCGGCGACCTCCGACCTGGGGTGGAGAGCCGAAAAGCTGAGAAATCCCTTAAGCTTCG 989
QY 5699 cctccatcgctgltccttactcttaacaataactccccaagccagtaglaagacgctta 5758
DB 990 CCTCATGCGCTGGTTCCTTACTTAACAATACTCCAGGCCAGTAGTAGTAAGCCCTTA 1049
QY 5759 tagaagctgaacccccaatagacctatataccttaactgctgattatgaacctata 5818
DB 1050 TAGACAGCTCGAACCCCATGACCTTATCCCTTACCTGAGCTATTAATTAACCTGATA 1109
QY 5819 cgggtgtactgtataatagcaactcgaagtgctgctccataagagcaccgtgagcctgaac 5878
DB 1110 GGGGTGTCACTGTAATATGACACTGAGGTGTGCTCCAGAGGCACTCGTGGCTTGAAC 1169
QY 5879 tgcattctgcctccgatgtataaccgcgtgttaanaagacacactcccaactagtc 5938
DB 1170 TGCAATTCCTGCTCGATTTGATTAACCCGCTGTTAAAGACACACTCCCACTAGTCC 1229
QY 5939 gtagttagtggttctatattgctgcccagagcaagaagaagaatactgtggggttcg 5998
DB 1230 GTAGTTATGAGTTCTATTGCTGCCAGGACAGAGAAAGAAATTAAGTGGGGTTCG 1289
QY 5999 ggaatccctctctagagatgagctgctgcaactcccaagatgagactggaatgagc 6058
DB 1290 GGAATCCCTTCTGTAGAGAGATGAGCTGCGCTCACTCCAAAGATGAGACTGGAATGCG 1349
QY 6059 cgatctctctccagacccggtaaatctcctctgtcaatctccgcccgggcaagtaca 6118
DB 1350 CGATCTCTCTCCAGACCCGGGTAAATTTCTCTTGTCAATTCGGCCGGCAAGTACA 1409
QY 6119 aaatgaagaactatataagaataagagctgctcccatagacttagatatactaaga 6178
DB 1410 AAGGATGATGAACATATATGAATAGAGCTGCTCCCATGAGACTTATGATTATCTAAGA 1469
QY 6179 taagtctactg--aaaggaanaacagaataatcaaaagtgtataatggtatgact 6236
DB 1470 TAACTTTCACCTGAAAAAGAAAAACAGSAAATATTCAAAAGTGTATATGATATGACT 1529
QY 6237 ggggaataagttttat-tatataggcggggagcaaggtccacttaaccatcgctta 6295
DB 1530 GGGGAATAGTTTTTATTAATATGCGGGGAGAGAGGCTCACTTAACCATTCGCTTAA 1589
QY 6296 ggaatagagacgggagacgaacccctgtgcaatgagacccgaataagtaactgctgagc 6355
DB 1590 GGAATGAGAGCGGGAGACGAACCCCTGTGTGCAAGTGGAGCCGATTAAGTATGCTGAGC 1649
QY 6356 agggggcccccggcctggaagcaccgcataactgtccgtgtgcccaatlaacctgcgc 6415
DB 1650 AGGGGCCCCGGCCCTGAGACCCGCGATTAATTGGCGGTGCCCAATTAACCTGCTGC 1709
QY 6416 ggcctgacataaacaagcgccctagcaacagtaacactgtgattgtattcctacaagc 6475
DB 1710 GGCGTGCATTAACACACCCGCTAGACAGCTACCACTGATTGATTCCTACCAACACCGC 1769
QY 6476 ctagaactcccaaggtgttctcgtttaagaagagagagagactctcaatctcaatcag 6535
DB 1770 CTGAAATCTCCCAAGGTTCTCTGTATGAAGACAGAGACACTTTCAGTCTCATCCAGG 1829
QY 6536 ggaacttcaagcatcaactccacgcactgtatgcactctctctctgtgtgctgtgc 6595
DB 1830 GAGCTTTCACCAAGCATCACTCACCGACCCGTATGCGCACTTCTTGTGTGCTTGTGC 1889
QY 6596 taccctcagggcctcttattatgaggggagtagctaaagaagaataatcaatgtgagca 6655
DB 1890 TATCTCAAGGCGCTCTTATTATGAGGGATGCTTAAAGAGAGAAATTAATGTGACCA 1949
QY 6656 aagagcatagaatacatgtlaactggggtcccgaaataagcttaaccctcaactgaagtt 6715

DB 1950 AAGGATATGAAATCAATGATACATGGGGTCCCGAAATAAGCTTACCTCAGTGAAGTT 2009
QY 6716 ccgggaaggggaaatgtatggaanaagctcccccatacccaacacactttgctaagta 6775
DB 2010 CCGGGAAGGGGACATGATAGAAAAGCTCCCATCCACCAACACTTGTGTAAGTA 2069
QY 6776 ctgtgttctagaagagcctcagaanaatcaagattacgtgttataaagtggt 6835
DB 2070 CTGTGATTATGAGCAGCGCTCAGAAATACAGATTTAGTACTGTTTAAACAGTGGT 2129
QY 6836 gggcagtgcaactggtgttaacccctgtgttccactcagtlctcaaccaatccaag 6895
DB 2130 GGGCATGCAATACTGGGTTTAAACCCCTGTGTTCACACTAGCTTCAACCAATCCAAAG 2189
QY 6896 attctgtgtcagtggtccaaatgtcccccagtagtcaactacactctgggaagtgtcc 6955
DB 2190 ATTCTGTGTCAATGCTCAATGCTCCCGAGTGTACATCCATCGTAGAAGTGTCC 2249
QY 6956 ttgatgaatatagactctgtataaccgaccaaagaagaacccgtatcccttaactag 7015
DB 2250 TTGATGAATATGACTCTGCTATTAACGACCAAAAGAGAACCCGTATCCCTTACCTAG 2309
QY 7016 ctgtatgctcgtatgaagagcgccgttggcgtgaagaaagagacgctgcccgtatca 7075
DB 2310 CTGTAAATGCTCGGATTAGGAGAGCGCGTGTGCGTAGAACAGGACAGCTGCCGTGATCA 2369
QY 7076 caagacacacagcagctgagagaagaagctgtgtgagctacaatgcgcatagacaagaatc 7135
DB 2370 CAGGACCAACGACGCTAGAGAAAGCACTGTGTGAGCTATATCGCCATGACAGAAATC 2429
QY 7136 tccgaagccttaagaagagctgttgaacaactgaagaagctcctgaactcttgtctgaag 7195
DB 2430 TCGGACCTTTAGAGAGAGCTGTGTAGCAACTTGAAGAGCTCTGACTTGTGTGAG 2489
QY 7196 tggttctacagaacccgaagggatagatctgctgttcttaagagaaggtggtatgtg 7255
DB 2490 TGCTTTACAGAACCGGAGGGGATTAGATCTGCTTTCTAAGAGAAGGTGGTATGTG 2549
QY 7256 cagccttaagaagaagaatgtctctctatagatcacatccatccagagacactcca 7315
DB 2550 CAGCTTTAAAGAGAAATGTTCTCTTATGATGACTCTGAGACCCCTTCAAGACTTCA 2609
QY 7316 tgaacaagcttgaanaaagttagagagagcgtcgaagggaaagaagctctacaggggt 7375
DB 2610 TGAGCAAGCTTAGAGAAAGTTAGAGAGCGCTGGAAGGAAAGAGAGCTGACAGGGGT 2669
QY 7376 ggtttgaagatgtgttcaacaggtctccttggaatgacaacccgtcttgcctgacgg 7435
DB 2670 GGTTTGAAGAGATGTTCAACAGGTCTCTTGTGATGACCACTGCTTCTGCTGACGG 2729
QY 7436 ggcctcagtagtctctgctccctgtacttaagttgggacctgtcttaataatggtttg 7495
DB 2730 GACCCCTAGTACTCTGCTCTCTGTTACTTACAGTTGGCCCTTGCTTAATTAATAGTTTG 2789
QY 7496 ttgcttgtttagagaaagatgagtgcaagtcagatataigtacttagaagaagtaacc 7555
DB 2790 TTGCTTTGTATGAGAACAGAGAGAGTGCAGACATCATGTACTTAAGCAACAGTAA 2849
QY 7556 aagcctcttgaagcacaagagaagaactgaacctcagccttccagttctaagaataact 7615
DB 2850 AAGGCTTCTGAGCCAAAGAGAACTGACCTCTACCTTCCCATTTCAATTAAGAACT 2909
QY 7616 attaaacagacagaagaagtgaggaaatgaagaatgaataatgaactaacctcccaaac 7675
DB 2910 ATTAAACAAGACAGAAGTGGGGAATGAAGAGATGAATAATGAACCTTAACCTCCAGAAC 2969
QY 7676 ccaaggaagttaataaagaactctaagtcccccgaaattacagaccctgtgctgcagat 7735
DB 2970 CCAGGAAGTTAATAAAGCTCTAAATGCCCCGGAATTCAGACACCTGCTGTGCTGACGT 3029
QY 7736 aaatagtagaaggtcacacttccatgtgtccagggcgtctatccctgagcaagtaag 7795


```
Db 3030 AATAGGTAGAGTGCACACTTCTATTGTTCCAGGGCTGCTATCCMGCCCTAAGTAAG 3089
QY 7796 ataacaggaaatgagtltagctaatcgtatcgtgallcttgraanaactgagcaccat 7855
    |||||
Db 3090 ATPACAGGAATGATTGACTAATGCTTATCTGATTTCTGTAACCGACTGCGACCAT 3149
QY 7856 agaagaattgattacacaltgacagccctagtgacctatctcaactgcaatctgcaactc 7915
    |||||
Db 3150 AGAAGAATTGATTACACATTGACAGCCCTAGTACCTATCTCACTGCATCTGTCACTC 3209
QY 7916 tgcccaaggagcccaagcaatgagagctccggagctatlttaaatgattgltccacag 7975
    |||||
Db 3210 TGCCCAAGAGAGCCACGAGATGCGAGCTCCGAGCTATTTAATAATGATTGTCACGCG 3269
QY 7976 agcgcgagctcgcgatatlttaaatgattgltccatgagagcgagctc-lcgataltc 8034
    |||||
Db 3270 AGCGCGGCTCTCGATATTTTAAATGATTGTTGCCACGAGCGGCGCTCTCGAATATT 3329
QY 8035 taaatlgattglttgtagcgaagagcttltgtgaacccataaaagctgtcccgat 8094
    |||||
Db 3330 TAAATGATTGTTGTGAGCGCACAGGCTTGTGTGAACCCCATTAATAAGCTGTCCCGAT 3389
QY 8095 tccgcactcggagccgcaatccctcctacccctcgtagtgaactgtgagcccaagcgc 8154
    |||||
Db 3390 TCCGCACTCGGGCGCGAGTCTTAACCCCTGCTGTGTAGACTGTGGGCCCCGACGCG 3449
QY 8155 gcttgaataaaaaactcctctgctgttgcatc 8187
    |||||
Db 3450 GCTTGGATAAAAAATCCTCTGTGCTTGGCATC 3482
```

Search completed: February 24, 2002, 07:30:19
Job time: 28334 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:30:19 ; Search time 10173.5 Seconds

(without alignments)
4793.420 Million cell updates/sec

Title: US-09-171-553B-9

Perfect score: 2956

Sequence: 1 tgccttttagggtagaac.....aaaaaaaaaaaaaaaaaaaaa 2956

Scoring table: OLIGO NWC
Gapop 60.0 , Gapext 60.0

Searched: 147140 seqs, 8248589755 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: GenEmbl:

- 1: gb_da:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_da:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_ov:*
- 21: em_or:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hgo_hum:*
- 31: em_hgo_inv:*
- 32: em_hgo_rtd:*
- 33: em_hgt_hum:*
- 34: em_hgt_inv:*
- 35: em_hgt_rtd:*
- 36: em_hgt_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8
SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2191	74.1	8209	6	A66553	A66553 Sequence 3
2	2089	70.7	8196	6	A66552	A66552 Sequence 2
3	1872	63.3	1974	6	A66559	A66559 Sequence 9
4	1763	59.6	8750	14	PEN133818	AJ133818 Porcine e
5	1743	59.0	3482	6	AX002804	AX002804 Sequence
6	1743	59.0	3482	14	PEN17013	Y12239 Porcine end
7	1654	56.0	7808	14	PEN13816	AJ13816 Porcine end
8	1636	55.3	8518	14	PEN13816	AJ13816 Porcine end
9	1572	53.2	8763	14	PEN293657	AJ293657 Porcine e
10	1499	50.7	1971	14	AF014162	AF014162 Porcine e
11	813	27.5	7333	6	AF038601	AF038601 Sus scrofa
12	813	27.5	7333	6	AR130474	AR130474 Sequence
13	626	21.2	8849	14	PEN133817	AJ133817 Porcine e
14	457	15.5	2462	6	AX002802	AX002802 Sequence
15	457	15.5	2462	14	PERENV1	Y12238 Porcine end
16	446	15.1	1042	14	PERY17012	Y17012 Porcine end
17	412	13.9	2000	14	PEN288588	AJ288588 Porcine e
18	375	12.7	3025	4	AF147808	AF147808 Sus scrofa
19	347	11.7	1968	14	PEN288589	AJ288589 Porcine e
20	321	10.9	746	14	PEN298073	AJ298073 Porcine e
21	318	10.8	590	14	PEN298074	AJ298074 Porcine e
22	318	10.8	2000	14	PEN288592	AJ288592 Porcine e
23	143	4.8	3320	6	A66551	A66551 Sequence 1
24	143	4.8	8918	14	PEN293656	AJ293656 Porcine e
25	139	4.7	2002	14	PEN288591	AJ288591 Porcine e
26	129	4.4	1996	14	PEN288590	AJ288590 Porcine e
27	116	3.9	1999	14	PEN288587	AJ288587 Porcine e
28	98	3.3	8132	6	AF038600	AF038600 Sus scrofa
29	98	3.3	8132	6	AR130475	AR130475 Sequence
30	91	3.1	633	6	AX052622	AX052622 Sequence
31	91	3.1	704	6	AX052621	AX052621 Sequence
32	91	3.1	4918	6	AX052637	AX052637 Sequence
33	91	3.1	6076	6	AX052636	AX052636 Sequence
34	91	3.1	6076	6	AX052647	AX052647 Sequence
35	91	3.1	7362	6	AX052634	AX052634 Sequence
36	91	3.1	7873	6	AX052638	AX052638 Sequence
37	91	3.1	8060	6	AR130473	AR130473 Sequence
38	91	3.1	8132	4	AF038599	AF038599 Sus scrofa
39	88	3.0	2008	14	PEN288586	AJ288586 Porcine e
40	74	2.5	4402	6	AX052635	AX052635 Sequence
41	62	2.1	1996	14	PEN288584	AJ288584 Porcine e
42	57	1.9	610	4	AF286463	AF286463 Sus scrofa
43	53	1.8	595	14	PEN298075	AJ298075 Porcine e
44	41	1.4	1591	10	AF161307	AF161307 Mus caroli
45	41	1.4	195423	2	AC087151	AC087151 Mus muscu

ALIGNMENTS

RESULT	1	A66553	8209 bp	DNA	PAT	29-MAR-1999
LOCUS	A66553	Sequence 3 from Patent WO9740167.				
DEFINITION	A66553					
ACCESSION	A66553.1	GI:4538106				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						

porcine endogenous retrovirus.
porcine endogenous retrovirus
Viruses; Retroid viruses; Retroviridae; Mammalian type C
retroviruses; 1 Mammalian type C virus group.
Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
PORCINE RETROVIRUS
Patent: WO 9740167-A 3 30-OCT-1997;
Q ONE BIORECH LTD (GB)
Location/Qualifiers
1..8209
/organism="porcine endogenous retrovirus"
/db_xref="taxon:61673"

CS 588..2162
 /codon_start=1
 /product="VIRAL CORE PROTEIN"
 /protein_id="CAB39347.1"
 /db_xref="GI:4538107"
 /translation="MGQVTPPLSLTLDHWEVSRANLSVQKKGPWQFCASEMP
 TDPVMSSEGTENSEIILAVKAIIFQOTGSHDPDEYILTMQDLAEDPPVAVKFWLN
 KPRKPGRIILALGKKNHSAKEVSSSVLPDRGADLAGTPTCSPTLSGCCG
 TSAPGAPVPEAGTASRGATPERTDEIALPLRTYRPGGLOLPQWPFSS
 ADLYNMTNHPPESEDPORGLTGVESIMSHOPTMDQOOLLOTLPTEPERILL
 RNKVPADGRPTOLONEIDMGFLTRPGDYTAGRESIKTYIRALVAGLGAARRP
 TNLAKEVWQNEPPEPSVFLRLMEAFRRFPEDPSAQAAVLAETGSAIDIR
 KTLAAREVQGEALDLVAREKYYRRETEEKQKREKEERDEERDSEENLT
 KTLAAVEGSSSRERERDFRIRSGRQSGNLGNTPLDKDQCAVCKEKGHWARNCPK
 KGNKGPVLALEDND"
 <2163..5747
 /codon_start=1
 /product="REVERSE TRANSCRIPTASE"
 /protein_id="CAB39348.1"
 /db_xref="GI:4538108"
 /translation="GRGSDPLPEPRVTLVGQPEVFLVDGAEHSLVLOPLGKKE
 KSMWVGATGQROYPWTRRTVDIGVGRTHSEIYIPGPVPLLRDILTKGAQISF
 BQGRPEVSNKKRTYVITLQIDDEVRLSPYKPPRODIOHLEQFQMAETAGGLA
 KQVPEVIOLKASATPEVSVQYPLSREAREGTPWHPQRLIQGLLVQSPWNTPLP
 VRKPGINDYRPVQDLREVNKRVQDIHPVNDYNLISALPPERMVTVLIDAEFCL
 RLHPTGQPLFAFEMRDPGTGRTGOLTWRLPOGFNSPTIDEDALHRLANRDIHQ
 VTLLQYVDDLLAGATKODLEGKALLLELSDLGABAKAOICRREVTYLGSLR
 GQGRMTLEARKKTYVOIRPAPTAKOYRELTGATGCRUMIPGATLALVPLTEKGS
 GRSWAPHEOKADAIKALLSAPALAPVPTKFLYDERKGVAVITQVLGMR
 PAVYLSKRLDPVAGSWPCLKATAVAAILVKADKLTLSQNTITVAPHALENTVQPP
 DRMTNARMTYQSLTLTERTVFAPPAALNPATLLPEETDEVTIDHQLIEETGVR
 KDLTIDPLTGEVLTWFTDSSYVVEGRNAGAAYVDGTITWASSLPETSQKQELM
 ALTOALRLAEKGSINITYDSRYAFATAYHAGAIYKQGLTISAGREIKKEILSLE
 ALHPEKRLAIHCPGHOKAKDLISRGNOADVAQAOAVLULPIIEPKAPEPRO
 YTEDEMOELKIDQSEPEPTCYGSEYELLPHKEGLEYVOHRLTHGTGKHQOL
 VRTSPYHVLRLPGVADSVYKPCPOLYANPSTRIPGKRLGSHPGANWEDTIEVK
 PAKYGNKYLVLEVPDTSVVEAVIPKERTSTVAKLIEEIPRFGIPISGDSGA
 FVAQVQSLAKILIGIDMKLHCAYRQSSQVERMNTIETLKLTEGGLANMALL
 PEVLFRVMTPGQGLPYELLYGGPPALFAISAVLLELSDPLSLKLEVRQ
 RAWKOLREAYSGDLOVPHRFQVDSVYRRRAGLLETRMKGPVILVLTPTAVAKE
 GTPWTHASHAVKPARPPDGSWAKETENLKLRLHRLVYNNNSPGQ"
 5620..7590
 /codon_start=1
 /product="VIRAL ENVELOPE PROTEIN"
 /protein_id="CAB39349.1"
 /db_xref="GI:4538109"
 /translation="MHPILSRHLPTRRGGEPRKRLRPLSFASIAMFLTLTPQASK
 RLIDSNPHRPLSTMLTIDPDGTVNSTRGVAAPGCTWMPHLCIRLINPAVSTP
 PNLVSTGYCCGTEKEKTCGSGSEFRRNSCVTSNDGKMPISLDRKFEFVN
 SGPGRYKMKLYKDKSCSPSDLYLKIISTEKTKGYSKVDWYELGNSFLLYGGAG
 SLLIRLRIETGTEPPVAMPDPKVAEOPALPEPHNLVQULSLRDLITQPSNS
 TTGLIPTNPNRSGVPKRGOLFSLIGAFQAIINSTPDITSSCWLCSGPPYE
 GMAKERFVYTKHNRQCTGSRNKLTLIEVSGKICDKAPPSHOHLSTVYVEOA
 SENQYLVPGYNMWAACNGLTPOVSVSNOSKRDVQVAPRYVYHEEVLDEYD
 YRNRPKREPVSLLTAVMLGLTAVGVTGTAALITTPQOULEKGEELHAANTEIDRA
 LKESVNLDESLSLSEVVLQNRGLDLFLLEGGLCAALKKECCGYVDHSAITDSM
 NKLRLKERRRERREADQGFEGFNRSFMTTLTSLALGPLVILLTLTVPCILNR
 FVAFVERVSAVOINVLROOYGLISQGETDL"
 BASE COUNT 2168 a 2064 c 2154 g 1823 t
 ORIGIN

Query Match 74.1%; Score 2191; DB 6; Length 8209;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2741; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 208 gagactgcaagtcacatcgcttcaagttggagattcaagctatgttagagccacc 267
 |||
 DB 5464 GAGACTTGCAAGTCCACATCGCTTCCAAAGTTGAGATTCAGTCTAGTAGACGCCACC 5523
 |||
 QY 268 gtgcaggaacacgcagactcggtggaaggacattaccgtaactttgacacaccaa 327
 |||
 DB 5524 GTGCAGGAACCTCGAAGCTCGGTGGAAGGACCTTATCTCGTACTTTTACACACACCA 5583

QY 328 cggctgtgaagaatcgaaagaaatctccacctgatacgaatcccaagcttaagctgagc 387
 |||
 DB 5584 CGGCTGTGAAGATCGAAGGAATCCCCACCTGGATCCATGATCCACAGCTTAAGCGCGGC 5643
 |||
 QY 388 caactcccgactcggggtggaagagccgaagaagacttgaaatcccttaagcttgcctcc 447
 |||
 DB 5644 CACCTCCCGACCTCGGGGTGGAAGCGGAAAGACGAAATCCCTTAAAGCTTGGCCCTCC 5703
 |||
 QY 448 atcgactggttcttactactaaacaaataactcccgagccagtagtaaacgcttaagac 507
 |||
 DB 5704 ATCGCTGGTCTCTTACTCTTAACAATACTCCCGAGCCAGTAGTAACGCCCTTATGAC 5763
 |||
 QY 508 agctgaaccccccatagacattatcccttacccttgctgattatgacccctgaagaggt 567
 |||
 DB 5764 ACTCGAACCCCATAGACCTTATCCCTTACCTGGTGGATATGACCTGATACGGGT 5823
 |||
 QY 568 gtcaactgtaataagcaactcgaggtgtgtcctctagaagcaacttgctgagcaatgat 627
 |||
 DB 5824 GTCACTGTAAATAGACATCGAGGTGTGCTCTTAGAGGACCTGTGTGCTGAACTGCAT 5883
 |||
 QY 628 tctctgctccgattgattaaacccgctgtttaaagcacactcccaactagtcgtagt 687
 |||
 DB 5884 TTCTGCTCCGATTGATTAAACCCCTGTAAAGACACACTCCCAACTAGTCCGTAGT 5943
 |||
 QY 688 tatgggtctatctgtcccaaggcaagaagaagaataactgtgagggttcttgggaa 747
 |||
 DB 5944 TATGGGTTTATGTCGCCACAGCACAGAAAGAAFAACTGCGGGTTCTGGGAA 6003
 |||
 QY 748 tccctctgtaggaatgtagctcgctgaactcccaagctgtagaacttgagcagatc 807
 |||
 DB 6004 TCCCTCTGTAGAGATGAGCTGCTGTCACTCAACGATGAGACTGGAAATGCGCGATC 6063
 |||
 QY 808 tctctccagagcgggtlaaatacttcccttgcattcctcggccggggaagtaacaaatg 867
 |||
 DB 6064 TCTCTCGAGACCGGGTAAATTTCTCTTTGCAATTCGGGCCGCGGCAAGTACAAATG 6123
 |||
 QY 868 atgaaactataaagaataagactgtctccccaatcaagacttagattatctaagaatg 927
 |||
 DB 6124 ATGAAACTATATAAGATTAAGAGTGTCTCCCATCAGACTTAAGATTATTAAGATAGT 6183
 |||
 QY 928 ttcactgaaagaaagaaacagaaataatctaaagttagataatgataagctgggga 987
 |||
 DB 6184 TTCACCTG--AAAGGAAACAGGAAATATTCAAAGTGGATTAAGTATGATGAGCTGGGGA 6241
 |||
 QY 988 atagttttttaaataatgtagcgaggagagaggtccactttaaaccatcgacttagata 1047
 |||
 DB 6242 ATAGTTTTTT-ATTATATGCGCGGGAGAGAGGTCCACTTAAACATTGCGCTTAGATA 6300
 |||
 QY 1048 gagaacgggagacgaacccccctgtggaatgggaacccgataaagtaactgtgctgaacagg 1107
 |||
 DB 6301 GAGACGGGACAGAACCCCTGTGCAATGGACCCGATTAAGTACTGCTGAAACAGGGG 6360
 |||
 QY 1108 ccccccggccttgaagcaacacgataacttgcgcgtgccccaaatlaactcgcttggagct 1167
 |||
 DB 6361 CCCCCGCGCTTGAGACACCGCATTACTTGCCGGTCCCAATTAACCTCGCTGGGCT 6420
 |||
 QY 1168 gacataacacgaacgcttagacaaagtaacacatgtagatttactctacaacacgcttaga 1227
 |||
 DB 6421 GACATTAACACGCGCGCTTAGCAACAGTACCACTGATTTGATTCCTAACCAACAGCCTAGA 6480
 |||
 QY 1228 aactcccaaggtgttccctgtttagaagagagagaaacttcaagtctcaacagggagct 1287
 |||
 DB 6481 AACTCCCCAGGTGCTCTCTTAAAGACAGACACTTCAAGTCACTATCCAGGAGACT 6540
 |||
 QY 1288 ttcocagacataaactccacacgaccttagacatgacattcttctgttggcttgcctacc 1347
 |||
 DB 6541 TTCAGAGCATTAACCTCAACGACCTGATGCACTTCTTCTTGTGGCTTTGTATACC 6600
 |||
 QY 1348 tcaaggctcccttatatagaggagatgagctaaagaagaanaatccaatgtgaccaaagag 1407
 |||
 DB 6601 TCAGGCGCTCTTATTATGAGGGGATGGCTAAGAAAGAAATTCATGATGACCAAGAG 6660

QY 1408 catagaatcaatgacatgagggggtcccgaaatgaagctaacctcgaagttccggg 1467
 |||||
 Db 6661 CATAGAAATCATGTACATGAGGGGGTCCGAAATAGCTTACCTCAGTAAGATTCCGGG 6720
 QY 1468 aagggagacatgacatgagaaagctcccccaccccaaacaccttgctatagatctg 1527
 |||||
 Db 6721 AAGGGACATGACATAGGAAAGCTCCCATCCACCAACACCTTGGCTATAGTACTGG 6780
 QY 1528 gttatgagcagcctcagaanaatcagtaattagtaacctggtatataacagtggtgagca 1587
 |||||
 Db 6781 GTTTATGACGAGCCCTCAGAAATACGATTTAGTACCTCGTTATTAACAGTGGTGGCA 6840
 QY 1588 tgcatactggtgtlaacccccctggtgtccacctcagttcctaaccaatccaagattg 1647
 |||||
 Db 6841 TGCATATAGGTTTACCCCTGTGTTCACCTCAGTTCACCAATCCAAAGATTTTC 6900
 QY 1648 tggatcaatgacaaatggtcccccagtggttacctcagtcagtgaggaatggtcctgat 1707
 |||||
 Db 6901 TGTGTATGCTCAAAATGCTCCCGAGTGTACTACCATCTGAGGAATGGTCTTGTAT 6960
 QY 1708 gaatagaactcgtatataaccgacaaagagaacccgtalcccttaacctagctgta 1767
 |||||
 Db 6961 GAATATGACATATGATATACGACCAAAAGAGAACCCGTATCCCTTACCTAGCTGTA 7020
 QY 1768 atgctcggatlaaggaacggtggtggtgagaaacggagacggtgcttgatcagaga 1827
 |||||
 Db 7021 ATGCTCGATTTAGGAGCGGCGCTTGCGTGAAGAACGAGCACTCCCTGATCAGAGA 7080
 QY 1828 cccagagagctagaagaagagactgtgtgagctacatgagcgcaatgacagaatcccgaa 1887
 |||||
 Db 7081 CCACGACGCTAGAGAAAGACCTTGGTGAACCTACATGCGGCCCATGACAGAGATCTCCGA 7140
 QY 1888 gccctaaaggagctgtgtgacaaactagaagagtcctcgtaccttctgtcgaatggt 1947
 |||||
 Db 7141 GCCTTAAGAGAGTCTGTACCAACTAGAGAGTCCCTGACTTCTTGTGTGAAGTGGT 7200
 QY 1948 ctacagacggggaggtatagatctgtgttcttaagaagaagtggtatgtgagcc 2007
 |||||
 Db 7201 CTACAGAACCGGAGGAGTATGATCTGCTTTCTTAGAGAAAGTGGTATATGTGACGCC 7260
 QY 2008 ttaaaaaagaatgtgtctctatgtatgacactcagagagccatcagagactccatgaa 2067
 |||||
 Db 7261 TTTAAAGAGAGATGTCTCTTATGATGATCAGAGAGCCATCAGAGACTCATGTAAC 7320
 QY 2068 aagcttagaaaaagttagagagggcgctcgaagggaaagagagtgctccgggggtgtt 2127
 |||||
 Db 7321 AAGCTTAGAAAAAGTTAGAGAGCGCTGCGAAGGAGAAAGAGCGCTGACCGAGGTGTTT 7380
 QY 2128 gaagagatggttcaacaggtctcctgtgatgacacacccgtcttctgtcgaagggccc 2187
 |||||
 Db 7381 GAAGGATGCTCAACAGGTCTCTTGTGATGACCACTGCTTCTGTGACGCGGCC 7440
 QY 2188 ctatagctcgtcctgtgttactactagttgagccttgctgttaataagttgtgtgc 2247
 |||||
 Db 7441 CTAGTAGCTCGTCTCTGCTTACTTACAGATTGGCCTTGCTTAATTAATAGGTTGTGCC 7500
 QY 2248 ttgtttagagaacgagtgagtgagctccagatcagtgactaggaacagttacaaagcc 2307
 |||||
 Db 7501 TTTGTTAGAGAACGAGTAGAGTGCAGTCCAGATCATGAGTCTTAGGACACATACCAAGGC 7560
 QY 2308 ctctcagcacaaggaagaaacagactcagcttccaggttctaagattagaactatga 2367
 |||||
 Db 7561 CTTTGACGCCAAGAGAGAACTGACCTTAGCTTCCAGTTCTTAAGATTAGACTATTTAA 7620
 QY 2368 caagaagaagtggtggaatgaagaatgaatgaatgaatgaatgaatgaatgaatgaatga 2427
 |||||
 Db 7621 CAAAGCAAGAGTGGGAGATGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG 7680
 QY 2428 aagtaataaagaagcttcaaatgccccgaattacagaacccgtgctgctgcaatgaata 2487
 |||||
 Db 7681 AAGTTAATAAAGAGCTTAAATGCCCCGGAATTAAGAGACCTGCTGGTGGCAGTAATA 7740
 QY 2488 ggtagaaggtacacttccatgtgttccagggcgtgctatccctgagtaagtaagataac 2547

Db 7741 GGTGAAAGGTACACCTTCTTATTTTCCAGGGCTGCTATTCGCGCTAGATGAATTAAC 7800
 |||||
 QY 2548 aggaatgagtgactaactacgtctatctgtgacttctgtataaacatgacgtgagcaataga 2607
 |||||
 Db 7801 AGGAATGAGTGTGACTATATGCTTATCTGATGATGTTGTTAAACCTGACGCACTAGAAG 7860
 QY 2608 aattgataacacattgacagccctcagtagtactatcactcaactgcaatctgcaactgcc 2667
 |||||
 Db 7861 AATTGATTACACATTGACAGCCCTAGTACCTATCTCAACATGCAATCTGTCACCTGCC 7920
 QY 2668 agagccacagcagatgagcagcctccgagagctatttaaaatgattgtgccaagagcgc 2727
 |||||
 Db 7921 AGGAGCCACGCAATGAGCGGACCTCCGAGCTATTTTAAATGATTTGCTCACGAGACGC 7980
 QY 2728 gggctctgataatttaaaatgattgtgctcagtgagcgggctcagatatttaaat 2787
 |||||
 Db 7981 GGGCTCTGATATTTTAAATGATTTGTCATGAGAGCGGCGCTCTGATATTTTAAAT 8040
 QY 2788 gatggtgtgagcagacagcgttgtgtgtaaccccaataaagctgtccgattccgca 2847
 |||||
 Db 8041 GATTGTTTGTGACGACAGCGCTTGTGTGAACCCCATTAAGCTGTCCGATTCCGCA 8100
 QY 2848 ctgagggcgcagagctcctcctacccctggtgtgtgtaacgtgtggtgccccagcggtg 2907
 |||||
 Db 8101 CTGGGGCGCGAGTCTCTTACCCCTGCGTGTGTCGACTGTGGCCACGCGCTTGG 8160
 QY 2908 aataaaatccctcgt 2956
 |||||
 Db 8161 AATTAATAATCTTGT 8209

RESULT 2
 A66552 8196 bp DNA PAT 29-MAR-1999
 LOCUS A66552
 DEFINITION Sequence 2 from Patent WO9740167.
 ACCESSION A66552
 VERSION A66552.1 GI:4538105
 KEYWORDS
 SOURCE
 ORGANISM
 porcine endogenous retrovirus.
 Porcine endogenous retrovirus
 Viruses, Retroviral viruses: Retroviridae; Mammalian type C
 retroviruses; 1-Mammalian type C virus group.
 REFERENCE
 1 (bases 1 to 8196)
 Galbraith, D.N., Haworth, C., Lees, G.M. and Smith, K.T.
 PORCINE RETROVIRUS
 Patent: WO 9740167-A 2 30-OCT-1997;
 JOURNAL Q ONE BIOTECH LTD (GB)
 FEATURES
 source
 1..8196
 /organism="porcine endogenous retrovirus"
 /db_xref="taxon:61673"
 BASE COUNT 2165 a 2061 c 2146 g 1820 t 4 others
 ORIGIN

Query Match 70.7%; Score 2089; DB 6; Length 8196;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 2739; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

QY 208 gaaactgtcgaagttccacatcgcttccaaatgtgagatcagctatgtttagagcagcc 267
 |||||
 Db 5450 GAGACTGTGCAAGTTCCACATCGCTTCCAAAGTTGAGATTCATGATTTAGAGCGCAC 5509
 QY 268 gtgaggaacacccagactcgggtggaagagacctatcgtactttgacacacaa 327
 |||||
 Db 5510 GTGAGGAAACCTCGAAGCTCGGTGGAAGGAGACTTATCTGTAATTTGACACACAA 5569
 QY 328 cggctgtgaaagtcgaagatctccacccgtggtcagatccacagttgaagctggcgc 387
 |||||
 Db 5570 CGGCTGTGAAGTGCAGGAAGATCCCACTGAGATCCATGATCCACGTTAAAGCTGGCG 5629
 QY 388 cactcccgctcgggttgagagccgaaagactgagaatcccttaagcttcgctcc 447

Db 5630 CACCTCCGACTCGGGGGAGAGCCGAAAGACTGAGAAATCCCTTAAGCTTCCCTCC 5689
QY 448 atcgcttgcttactactactactactactactactactactactactactactactact 507
Db 5690 ATGCGCTGCTCTTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 5749
QY 508 agctcgaaeccccataagacttactactactactactactactactactactactact 567
Db 5750 AGCTCGAAGCCCCATFAGACCTTTATCCCTTACCTGCTGATTTATGACCTGATACGGGT 5809
QY 568 gtaactataaagactactactactactactactactactactactactactactactact 627
Db 5810 GTCACTGTAAATACACTCGAGGTGCTCTCTAGAGGACCTGCTGCTGAACTGCAAT 5869
QY 628 tctcgctcgatgtatataaccccgctgttaaaagacacccctcccaactcagtcagat 687
Db 5870 TTCTGCTCCGATGTAAACCCCGCTGTAAATAGCCACACTCCCAACTGATCCGAT 5929
QY 688 tatgggtctatgtctgcccagacacagagaagaataactgtggggttctgggaa 747
Db 5930 TATGGGTTCTATTGCTGCTCCAGGACAGAGAAAGAAATAGTGGGGTCTGGGAA 5989
QY 748 tctctctgtagagatgagctgctcaactccaagatggagactggaatgctggatc 807
Db 5990 TCTTCTGTAGAGATGAGACTCGTCACTCCACGATGAGACTGGAAATGGCCGATC 6049
QY 808 tctctcgaagacccgggtataatctctcttgcatactccggccgggcaagtaaaatg 867
Db 6050 TCTCTCCAGACCGGGTAAATCTCTTGTCAATTCGGCCGGGCAAGTACAAATG 6109
QY 868 atgaactataaagataagagctgctcccaactcagactgtgataactaagaatg 927
Db 6110 ATCAAACTATATAAATAGAGACTCTCCCTCACTGACTGTAGATTATTAAGATTAAT 6169
QY 928 ttaactgaaaaaagaaacaggaataatcaaaagtgtgataatgtaagtgctgggga 987
Db 6170 TTCTACTG- AAAGGAAAGACGAAATATTCAAAAGTGTATATGATGAGCTGGGGA 6227
QY 988 atagttttaataatagctgggggagagaggtcccaacttaaccatcgcttagata 1047
Db 6228 ATATGTTTTT-ATTATATGCGGGGAGAGAGGCTCCACTTAACTTAACTTGCCTTAGATA 6286
QY 1048 gagaacgggagacagacccccctgtgcaatggagaccggaataaagtactgtgtaaaagg 1107
Db 6287 GAGACGGGAGACGAACCCCTCTGGCAATGAGAACCCGATTAAGTACTGACTGAACAGGG 6346
QY 1108 cccccggccctggagccacgcataactgcccgtgcccacatlaacctgctggcct 1167
Db 6347 CCCCCGGCCCTGGAGCCACCCGATTAATGCGGTGCCCCAATTAACTTCGCTGGGCT 6406
QY 1168 gacataacacagccctagcaacagtagtaccgtgattgatttcttaaccaacagcctaga 1227
Db 6407 GACATTAACACAGCCCTCTAGCAACAGTACCACTGATGATTCTTAACCAACAGCCGTAAGA 6466
QY 1228 aactccccaggtgtctctgttaagacagagacagactctcagltcatcatcaggagct 1287
Db 6467 AACTCCCCAGGTTCTCTGTAAAGACAGACAGAGACTCTCAAGTCACTATCCAGAGACT 6526
QY 1288 ttccaagccatcaactccacgcagacctgtatgccaacttctgtgtgctgtctatcc 1347
Db 6527 TTCCAAGCCATCACTCCACGACCCCTGATGCTCACTCTTCTTGTGCTTGTGCTTTC 6586
QY 1348 tcaaggctctcttattagagggatggcttaagaagaataatcaatgtgacaaagag 1407
Db 6587 TCAGGGCTCTTATTATGAGAGGATGCTTAAGAAAGAAATCAATGTGACCAAGAG 6646
QY 1408 catgaataatcaatgtacatgggggtcccgaaataagcttaacctcactgaatgctggg 1467
Db 6647 CATGAATCAATGATACATGGGGTCCCGAAATAGCTTACCTCACTGAAGTTTCCGGG 6706
QY 1468 aaggggacatgcatagaaaaagctccccatccccacaacactgtgctatagactgtg 1527
Db 6707 AAGGAGATGATAGTAAAGAAAGCTCCCATCCCATCAACACTTGTATAGTACTGTG 6766
QY 1528 gttatgagcagggcctcagaanaatcaglattaglaactcgtgtataaagctgtgggga 1587
Db 6767 GTTATAGACAGGCTCAGAAATCAGTATTAGTACCTCGGTTATATACAGGTGGGGA 6826
QY 1588 tgcataactgggttaacccccctgtgttccacctgaagttctcaaccaatccaagaatgtg 1647
Db 6827 TGCATATGAGGTTAAACCCCTGTGTTCACCTAGCTTTCACCAATCCAAATATTC 6886
QY 1648 tgtglatgtccaaatcgtcccccagtgtaactaccatcccgaggaaagttgctctat 1707
Db 6887 TGTGATGATGTCOAATGCTCCCGAGGTACTACATCTCGAGAGAGTGTCTTGTAT 6946
QY 1708 gaatatgactatcgylataacccgacaaaagaacccglatcccttaacctagctgt 1767
Db 6947 GAATATGATTAATCGGTTATTAACCGACCAAAAGAAACCGTATCCCTTAACCTACTGTA 7006
QY 1768 atgctcgatlagagagcggtggtgggtagaagaagagagagctgcccgtatacagga 1827
Db 7007 ATGCTCGATTTAGGACGGCCCTTGGCGTAGAACAGGACAGCTGCCCTGATACAGGA 7066
QY 1828 ccacagcagctagagaaagactgtgtgactatactgagccatgacagaaatctccga 1887
Db 7067 CCACAGCAGCTAGAGAAAGACTTGTGAGCTACATGCGGCAATGACAGAAATCTCGA 7126
QY 1888 gcttaagagagctgttagaacacttaagaagctcctgacttctgtcgaatgtgt 1947
Db 7127 GCCTTAAGAGATCTGTAGAACCTTAAGAAAGTCCGACTTCTTGTGTAAGTGTGT 7186
QY 1948 ctacagaacccggaggggattagatctgttcttaagaagaaggtgtgtgtgagcc 2007
Db 7187 CTACAGAACCGAGGGGATTTAGATCTGCTGTTCTTAAGAGAGTGGTATTGTGCAAGC 7246
QY 2008 ttaaaagaagaatgttcttctatgtgataactcaggaagccatcaagactccatgaac 2067
Db 7247 TTAAGAGAGATGTGCTTATATGATGATCACTCAGAGGCACTCAGACTCATGATAC 7306
QY 2068 aagcttaaaaaagttagaagaagcgtcgaagggaagaagaggtgcccaggggtgtt 2127
Db 7307 AAGCTTAAGAAAGATTATAGAGAGGCTGCAAGGGAAGAGAGGTGACAGGGGTGCTTT 7366
QY 2128 gaagatgtgtcaacaggtctccttgaatgacacccctgtcttctgtcgaagggccc 2187
Db 7367 GAAGATGTCTCAACAGAGTCTCTTGGATGACCACTGCTTCTGCTGAGAGGGGCC 7426
QY 2188 cttagtctcgtcctcgttactactacagttgggctgtgttaataagttgtgccc 2247
Db 7427 CTATAGTCTCCTGCTCTGTTACTTACAGTTGGGCTTGTAAATTAATAGTTGTGGCC 7486
QY 2248 ttgttagaagaagtgagtgcaatcagatcatgtttagttaggaagaagtaacgaagc 2307
Db 7487 TTTGTATAGAGAGAGAGTATGCACTGATGCACTATGATGATGATGATGATGATGATGAT 7546
QY 2308 ctcttgagccaagagagaactgaactcctcagccttccaggttcaagaatagaactttaa 2367
Db 7547 CTCTTGAGCCAGAGAGAACTGACCTTACCTTCCAGTTTAAAGATTAAAGACTATTAA 7606
QY 2368 caagacaagaagtgggaatgaaagatgaanaatgaactaaacctcccaagacccag 2427
Db 7607 CMAAGACAGAGAGTGGGAATGAAGAGATGAATAATGAACCTTAACCTCCCAAGACCCAG 7666
QY 2428 aagtataataaaagcttlaaatgcccccgaaatatacagaccctgtgtgctgcagtaata 2487
Db 7667 AAGTTATATAAAGCTCTTAATGCCCGCAATTMCAGACCTGCTGCTGCCAGTAATAA 7726
QY 2488 ggtagaagttcaacttctctatgttccagggcctgtctatcctgtgcttaagtaataac 2547
Db 7727 GGTAGAAGGTCAACTCTTATTTGTTCCAGGGCTGCTATCTGGCTTAAGTAAGTAAAC 7786
QY 2548 aggaataggttagactaacctcgttctgtatctgttaaaactgtgacacatagaag 2607
Db 7787 AGGAATGAGTTGACTTAATGCTTATCTGATTTGTAAACTGACTGGCACCATTAAGAG 7846

QY 2608 aattattacacattgacagccctagtgacctatctcaactgcaatctgtcactctgccc 2667
|||||
Db 7847 AATTGATTACACATTGACAGCCCTAGTACCATCTCAACGTGCAATCTGTCACTGTGCC 7906
QY 2668 aggaagccacagcagatgagaccccgagctattttaaatgattgltccagagagc 2727
|||||
Db 7907 AGGAGCCACGACATGCGAGCTCCGGAGGATTTTAAATGATTTGGTCCACGAGCGC 7966
QY 2728 gggctctgatatatttaaaatgattggtccatgagagcgagctctgattttaaat 2787
|||||
Db 7967 GGGCTCTCGATATTTTAAATGATTTGGTCCATGAGCGCGGCTCTGATATTTTAAAT 8026
QY 2788 gattgattgtaacagagcttltgtgaaccccaataaagctgtccgacttcgca 2847
|||||
Db 8027 GATTGTTTGTACGACGACAGGCTTTGTGTGAACCCCATTAAGCTGTCCGATTCGCA 8086
QY 2848 ctccgggcccagctcctctactccctgctgltgtactgactgtggcccgagcgcttg 2907
|||||
Db 8087 CTCGGGGCCGACGCTCTGACCCGCGTGGTGTACGACTGTGGCCCGGCGGCTTGG 8146
QY 2908 aataaaatcctctgctgttgcacaaaaaaagagagagagagagagagagagag 2956
|||||
Db 8147 AATAAAATCTCTGCTGTGTCATCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 8195

RESULT 3
A66559 1974 bp DNA PAT 29-MAR-1999
LOCUS A66559
DEFINITION Sequence 9 from Patent WO9740167.
ACCESSION A66559
VERSION A66559.1 GI:4538112
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1974)
AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
TITLE PORCINE RETROVIRUS
JOURNAL Patent: WO 9740167-A 9 30-OCT-1997;
Q ONE BIOTECH LTD (GB)
FEATURES
source 1..1974
location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 534 a 489 c 483 g 468 t
ORIGIN

Query Match 63.3%: Score 1872; DB 6; Length 1974;
Best Local Similarity 99.9%: Pred No. 0;
Matches 1972; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 364 atgcatcccaaglttaagctgagcagcctccgactcgggtgagagccgaaagactg 423
|||||
Db 1 ATGCAATCCCAAGTAACTGAGCGCCACCTCCGACTCGGGGTGAGAGCCGAAAGACTG 60
QY 424 agaatcccttaagctgctccatcgctgtgttcttaaccttaataactcccg 483
|||||
Db 61 AGAATCCCTTAAGCTTCCCTCCATCGCTGCTTCTTACTCTTAACATACTCCCGAG 120
QY 484 ggcagtagtaaacgcttataagactcgaaccccatagacacttaaccctactcg 543
|||||
Db 121 GCCAGTAGTAACGCTTATAGACACTGGAACCCCATAGACCTTATCCCTACTGCG 180
QY 544 ctgattattgacccctatacaggtgtcactgtaaatagacactcgaaggtgtcctctaga 603
|||||
Db 181 CTGATTATTGACCCCTGATAGCGGTGTCACTGTAAATAGCACTCGAGGTGTGCTCTAGA 240
QY 604 ggcacctggtgagcctgaacgacttctgctcccgatgatttaaccccgctttaaagc 663
|||||
Db 241 GGCACTGGGTGGCTGAAGTCAATTTCTGCTCCGATTATTAACCCCGCTGTTAAAGC 300

QY 664 acacctcccaacccagtcgtagtataaggttctatgtctgcccagagcaagagaag 723
|||||
Db 301 ACACCTCCCAACCTAGTCCGAGTATGAGTTCATTTGCTGCCAGGACAGAGAAAGAG 360
QY 724 aaatctggtgggttctgaggaaatcctctgtatagagagatgagagctgctacactcaac 783
|||||
Db 361 AAATCTGTGGGGTCTTGAGGAATCTTCTGTAGAGATGAGAGCTGCTCACCCTCCAAC 420
QY 784 gatgagactgaaatgagccgactcctccagagccgggtaaatctcttltcaat 843
|||||
Db 421 GATGAGAGCTGGAATAGCGGATCTCTCCAGAGCCGGGTAATAATCTCTTGTCAAT 480
QY 844 tccggccgggcaagatcaaaaatgaaactataaagaataagagctcctccatca 903
|||||
Db 481 TCCGGCCCGGCAAGTCAAAATGATGAATATATTAAGATGAAGCTCTCCCATCA 540
QY 904 gacttagattatccaaagataagtttactgaaaaaagaaacggaaatattcaaaag 963
|||||
Db 541 GACTTAGATTATCTAAAGATTAAGTTCACCTGAAGAAAGGAAACAGAAATATTCAAG 600
QY 964 tgaataatgtaagctgagagatagttttaaataatagtcggggagagaggttcc 1023
|||||
Db 601 TGGATTAATGCTATGAGCTGGGGAATAGTTTATTAATATGCGGGGAGGAGCGGTCC 660
QY 1024 acttaacacatcgctccttagatagacagggagacaaacccctgtgcaatgagacc 1083
|||||
Db 661 ACTTTAACCAATTCGCTTAGATAGAGACGGGACAGAACCCCTGTGGCAATGGGACC 720
QY 1084 gataaagctagctgtagaagagggcccccggccttgagagccacgactaactgaggtg 1143
|||||
Db 721 GATAAAGTACTGCTGAACAGAGGAGCCCGGCTCCGAGGACACGCAATATTCGCGTG 780
QY 1144 ccccaatcaactcgtcgtcggccctgacataacacagccgctagcaacagtaacctgga 1203
|||||
Db 781 CCCCAATTAACTGCTGCTGCGGCTGACATTAACAGCGCGCTAGCAACAGTACACTGGA 840
QY 1204 ttgatctcacaacagccttagaagactcccgaggttctctgtttaaagacagagagaga 1263
|||||
Db 841 TTGATTCCTTACCAACAGCGCTAGAACTCCCGAGGTTCCTGTTAAGACAGGACAAGA 900
QY 1264 ctcttaagctcctacagagagagcttccaaacacatacctccagccctgagtcact 1323
|||||
Db 901 CTCTTCACTCTCATCCAGGAGCTTTCCAMGCCATCACTCCAGGACCTGATGCCACT 960
QY 1324 tctctctgtgcttgcatacctcagagccctctatataagagggatgagctaaagaa 1383
|||||
Db 961 TCTTCTTGTGTGCTTGTGCTATCTCCTAGGCGCTCTTATTAAGGGGATGGCTAAAGAA 1020
QY 1384 agaaatcaatgtgaccaaagacatagaatcaatgtacataggggttccgaaataag 1443
|||||
Db 1021 AGAAATTCATATGTCACCAAGAGCACTAGAAATCAATGTACATGGGGGTCCCGAAATAG 1080
QY 1444 cttaaccctactgaagtttccgggaagggacatgcataggaagaagctcccccactccac 1503
|||||
Db 1081 CTTACCTCTCACTGAAGTTCGGAAGGGGACATGCAATGAGAAAGCTCCCCATCCAC 1140
QY 1504 caacaccttgctatagactgtgttataagagagcctcagaanaatcagatattagta 1563
|||||
Db 1141 CAACACCTTGTGCTATAGTACGTGGTTATAGGACAGGCTCAGAAATACATATTAGTA 1200
QY 1564 cctgttataacaggtgtgagcatgaatcctggtttaaaccctctgttccaactca 1623
|||||
Db 1201 CCTGGTTTAAACAGGTGTGGGCAATGCAATCTGGGTAAACCCCTGTGTTTCAACCTCA 1260
QY 1624 gtcttcaaccaaaccgaagattgtgttcaatgtgtccaaactcgtcccccaggttactac 1683
|||||
Db 1261 GTCTTCAACCAATCCAAAGATTTGTGTCTCATGTGTCCAAATCGTCCCGAGGTACTAC 1320
QY 1684 calccctgagaagtgctctgtatgataatgactatcgatlaaacggaacaaagagaa 1743
|||||
Db 1321 CATCTGAGGAAGTGTCTTGTGATGATATGATCATCGTAAACCGCACAAAGAGAA 1380
QY 1744 ccgctatcccttaccctagctgtaatgctcgatlaaggaagcgctgtggtgtaggaaca 1803

[illegible]

Query Match 59.6%; Score 1763; DB 14; Length 8750;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 2913; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 tgcctttaggttaggaacaccccttgagacagtttggtctgaacccctatgaattgctct 60
 |||||||
 Db 5736 TGGTTTAAAGGTTAGGAACACCCCTGGACAGTTGGGCTGACCCCTATGATGCTCT 5795

QY 61 acgggggagaccccccggttgtagaattgcttcgtacatagtgctatgtgctctt 120
 |||||||
 Db 5796 ACGGGGAGACCCCGGTTGGTAAGAAATGCTTCTGTATATGCTGATGCTGCTTT 5855

QY 121 cccagcccttgctctcctagacagcgcctgagtggttgaggaacagcgtaggaagc 180
 |||||||
 Db 5856 CCCAGCCTCTGTTCTTAGGCTCAAGGCTCGAGTGGGTGAGGACAGAGCTGGAAGC 5915

QY 181 agctcggagagcctactcagagagagagacttgcaagttccacatcgctccaaagt 240
 |||||||
 Db 5916 AGCTCCGGAGGCTACTACTGAGAGAGAGACTTGCACATTCACATCCCTCCAAAGTGG 5975

QY 241 gagattcagctctatgttagagcgcacgtgcaggaagacctcgagactcgtaggaagagac 300
 |||||||
 Db 5976 GAGATTGAGTCTATGTTAGACGCCACCTGACAGAAACCTCGAGACTCGTGGTGAAGGGCC 6035

QY 301 ctatctcgtacttttgaccacacacacagcgctgtgaagtcgaaggaalctccacctgga 360
 |||||||
 Db 6036 CTATCTCGTACTTTTACACACACCAAGGCTGTGAAGTTCGAAAGAAATCTCCACTGGA 6095

QY 361 tcaatgatccccagtagtaagcttgagccacgtccgagctcgaggttgaggaagccgaagaa 420
 |||||||
 Db 6096 TCCATGATCCCAAGCTTAACCTGAGCCACCTCCGACTCCGAGTGGAGTGAAGCCGAAAGA 6155

QY 421 ctgagaatccaccttaagcttcgctccatcgcttggttctcttaacttaacaaactcc 480
 |||||||
 Db 6156 CTGAGAAATCCCTTAACCTTGCGCTCCATCGCTGGTTCCTTACTTAACAATAACTCCG 6215

QY 481 caaggcagtagtaacgacctatagagacgtcgaaaccccatagaccttatccctacc 540
 |||||||
 Db 6216 CAGGCGAGTAAACGCTTATAGACAGCTCAACCCCAATGACCTTATCTCCCTTACC 6275

QY 541 tggctgattatgacccctgaagcggtgctacgttaaatgacactcgaggtgctgctct 600
 |||||||
 Db 6276 TGGCTGATTATTGACCTCGATACGGGTGCTCACTGTAAATAGCACTGAGAGTGTGTCTCC 6335

QY 601 agaggcaccttggtgctgaactgacatltctgctccgattgataacccgcgtgttaa 660
 |||||||
 Db 6336 AGAGGACCTGTGGCTGAACTGACTGATTTGCTCCGATTGATTAACCCCGCTGTATAA 6395

QY 661 agcaacctcccaacctagctcgaagttaaggttctatgtgcgacagacagagaagaa 720
 |||||||
 Db 6396 AGCAACACTCCCAACCTAGTCCGTATAGGTTCTATTGCTGCCAGGACACAGAGAAA 6455

QY 721 gagaataacttggtggtctcggggaaactcctctgttagagagatggagctgctcactcc 780
 |||||||
 Db 6456 GAGAAATGACTGTTGGGTTCTGAGAAATCTTCTGTAGAGATGAGACTCGTCACTCC 6515

QY 781 aacgattgagactggaataatgacgactctctccaggaacgggtgaataatccctgtgc 840
 |||||||
 Db 6516 AACGATGAGACTGGAATGAGCGGATCTCTCCAGGAGCGGTAATAATCTCTCTTTGTC 6575

QY 841 aattccggccgggcaagtacaaaatgaataacatataaagaataagagctgctcccca 900
 |||||||
 Db 6576 AATTCGGGCCGGGCAAGTGAACAAATGATTAAGATTAAGATGAGAGTCTGCCCA 6635

QY 901 tcaagacttagattatcaagaataagtttcaactgaaaaagaaacgggaaatattcaaa 960
 |||||||
 Db 6636 TCAGACTTATGATTATCTAAAGATTAAGTTCACTGAAAAAGAAACAGGAATAATTCAA 6695

QY 961 aatgataataatgataatgacgtgaggaatagttttaaataataatgacgggagagcag 1020
 |||||||
 Db 6696 AACTGATTAATGATGATGAGTGGGATAGTTTTTATTAATATATGCGGGGAGCAAGG 6755

QY 1021 tccacttaaccatcgctcgttagagtagagacggggacagaaacccctgtgccaatggga 1080
 |||||||
 Db 6756 TCCACTTTAAACATTTGCTTATGATGATGACAGGGGACAGAAACCCCTGTGGAGTGGGA 6815

QY 1081 ccgataaagtagctgctgaacagggcccccggcccttgagagcaacgacataactgcg 1140
 |||||||
 Db 6816 CCGGATTAAGTACTGGCTCAACAGGGGCCCCGGGCTTGAGACCCACCATTAATCTTGCGG 6875

QY 1141 gtgcccacatlaacctcgctcggtcgactgaacataacacgacgacctagaacagtaacct 1200
 |||||||
 Db 6876 GTGCCCCAATTAACCTCGCTGCGGCTTGACATTAACACAGCCGCTTGAAGCGTACACT 6935

QY 1201 ggaattatctcctcaacacgcacctagaacaccccaagtgcttcgtttaaagacaggaag 1260
 |||||||
 Db 6936 GGAATTGATCTTACCAACACGCTTGAATAACCTCCCAAGTGTTCGTTTAAGACAGAGAG 6995

QY 1261 agactctcagctcactcagagagacttccaaagcaactccacacgacccgtgaagc 1320
 |||||||
 Db 6996 AGACTCTCAGTCTCATCCAGAGGACTTTCAAAGCCATCACTCAACCGACCTGATAGCC 7055

QY 1321 actctctctgtggtcttgctatctcctaagggccctccatlatagaagggatggtctaa 1380
 |||||||
 Db 7056 ACTCTCTGTGCTTGTGCTTGTCTATCTCAAGGCTCTCTTATTATGAGGAGATGGCTAAA 7115

QY 1381 gaaagaatltcaatgtgaccaaagacalagaatltcaatlygaggtgccgaat 1440
 |||||||
 Db 7116 GAAGGAATAATTCATATGAGACCAAGAGCATAGAAATCAATGTCATGGGGGCTCCGAAT 7175

QY 1441 aagctaacctcaactgaagattccgggaagggagacatgatatagaaagctcccatcc 1500
 |||||||
 Db 7176 AAGCTTACCCCTCACTGAAAGTTTCCGGGAAGGAGCATGATGAGAAACCTCCCATCC 7235

QY 1501 caccacacacttgactatagctggtgtttagaacagccctcagaanaactcaatatta 1560
 |||||||
 Db 7236 CACCAACACCTTGTCTATGATAGTGTGTATGACAGGCTCAGAAATCAGATTATA 7295

QY 1561 gtaccgtgattataacaggtggtggtgacatgaactggtgttaacccccgtgtttccacc 1620
 |||||||
 Db 7296 GTACCTGTTATTAACAGGTGTGGGATGCAATACGTGGGTTAACCCCGTGTCTCTCC 7355

QY 1621 tcaagtctcaacaaatcaagaatttggtgtcatgtgtccaaatctgcccagtgatc 1680
 |||||||
 Db 7356 TCAGTCTTCAACCAATCCAAAGATTTCTGTCTATGATGTCCAAATCGTCCCGCAGGTAC 7415

QY 1681 taccatctcgaagaagtgtgctctgataatgataactcgtataacacgacccaaaga 1740
 |||||||
 Db 7416 TACCATCTTGAGGAAGTGTCTTGAATGATATGACTATGCTATTAACCGACCAAAAAA 7475

QY 1741 gaaccggtatcccttaacctagctgataatgctcggtatagggacggcgttgctgaaga 1800
 |||||||
 Db 7476 GAACCCGTATCCCTTACCCTTAGCTGATATGCTGATTTAGGAGGCGGCTTGGCGTAGGA 7535

QY 1801 acaaggagacgtgcccctgatcacaggaacacagacagcttagaagaagacttgtagcta 1860
 |||||||
 Db 7536 ACAGGAGACAGCTGCCCTGATCACAGAGACACAGACGCTAGAGAAAGACTTGTGAGCTA 7595

QY 1861 catcgccatgacagaagaatctccgagcttlaagaagctcgttlaacacactagaaga 1920
 |||||||
 Db 7596 CATCGGCCATGACAGGAAGATCTCCGAGCTTAAAGAGTCTGTTAAACACTTAGAGAG 7555

QY 1921 tcccgcactctctgtcgaagtggtctcaagaacggagggattagactgctgltt 1980
 |||||||
 Db 7656 TCCCTGACTCTTTTCTGGAAGTGTTCACAGAACCGAGAGGATTAAGATGCTGTTT 7715

QY 1981 ctaagaagaagtgtgtttagtgcagccttaaaagaagaatgtgtctctatgtatgacac 2040
 |||||||
 Db 7716 CTAAGAGAGGTGGGTATGTGACACCTTAAAGAAAGATGTTCTCTATGATGATCAC 7775

QY 2041 tcaagagcaataagagactcattagaacagcttgaagaagaagtttagaagagcgtcgaag 2100
 |||||||
 Db 7776 TCAGGAGCCATCAGAGACTCCATGAGCAACCTTACAGAAAGGTTTAGAGAGGCGTCAAGG 7835

QY 2101 gaagagagagctgaccaggggtggttgaagagatggttcaacaggtctccttgatgacc 2160
 DB 7836 GAAAGAGAGGCTGACACAGGGGTGTTTAAAGATGTTCAACAGAGTCTCTTGATGACC 7895
 QY 2161 accggtctctgctctgaagggcccttagtccgtccgttacttaacagttggg 2220
 DB 7896 ACCGCTCTTCTGCTCTACCGGACCCCTAGTACTCCGCTCCGTTACCTTAACAGTTGG 7955
 QY 2221 ccttgcttaataatagttgtgtccttltgtagaagacaggtgagtgacacagatc 2280
 DB 7956 CTTGCTTAATTAATAGTGTGTGCTTTGTGTAGAGACAGAGATGACAGTCCATC 8015
 QY 2281 atgtacttaggcaacagttacaaagcctcttgagccaagagaaactgaacttagct 2340
 DB 8016 ATGCTACTTAGGCAACAGTACCAAGGCTCTTGAGCCACAGAGAACTGACCTTACCT 8075
 QY 2341 tcccaagttctaagttagaactataacagacaagaagtgaggaaatgaaagatgaaa 2400
 DB 8076 TCCCAAGTTCTAGATTAGAACTATTAAACAAGACAGAAAGTGGGAATGAAAGATGAAA 8135
 QY 2401 tgcacactaacctccccaagaaacccaagaaatlaataaaagctctaagtccccgaatt 2460
 DB 8136 TGCACACTTAACCTCCACAGAACCCAGGAGATTAAATAAAAGCTTAATGCCCCCAATT 8195
 QY 2461 acaagacctgtgctgctgcaagtaaatagtagaagagtcacacttccatgttccagggc 2520
 DB 8196 CCAAGACCTGCTGCTGCTCCAGTAAATAGTAAAGGTCACACTTCTTATGTTCCAGGGC 8255
 QY 2521 ctgctatccctgctgaagtaagaacaagaagaatgagttgactaatcgtctatctgatt 2580
 DB 8256 CTGCTATCTGCTGCTGAAGTAAACAGAAATGAGTTGACTTAATGCTTATCTGATT 8315
 QY 2581 ctgttaaacctgactgagcacaatagaagaattgtatcacatltgacagccctagtgacct 2640
 DB 8316 CTGTAAACAGACTGCGACCATAGAAAGATTGATTACACTTGACAGCCCTAGTGACCTA 8375
 QY 2641 tccaactgcaatctgtaactgccccagagagcccaagatgagacctccagagacta 2700
 DB 8376 TCTCAACTGCAATCTGTCACTCTGCCAGAGAGCCCAAGAGATGCGAGCCTCCGGAGCTA 8435
 QY 2701 ttttaaatgattgtgtccagagagcgagctctcgatatttaaatgattgtgtccatg 2760
 DB 8436 TTTTAAATGATGTGTCCACAGGAGCGGGCTCTCGATTTTAAATATTGTGTCCACG 8495
 QY 2761 gagcgagagctctcgatattttaaataatgtgtgtgtgagcaagagcttgtgtgaa 2820
 DB 8496 GAGCGGAGGCTCTCGATTTTAAATGATGTGTGTGACGACACAGGCTTGTGTGAA 8555
 QY 2821 ccccaataaagctgtccagatccgaactcgagcgagcgagctcttaaccccggtgtg 2880
 DB 8556 CCCCAATAAAGCTGTCCGATTCGCCACTCGGCGCGCACTCTTACCCCTGCGTGTGG 8615
 QY 2881 taccagctgtggcccgagcgagctgtgaataaaatcctctgtgtgtgtgacaa 2936
 DB 8616 TACGACTGTGGGCCCCAGAGCGCTTGGAATAAAATVCTTGCTGTGTGATCA 8671

RESULT 5
 LOCUS AX002804 3482 bp DNA PAT 21-AUG-2000
 DEFINITION Sequence 3 from Patent WO9853104.
 ACCESSION AX002804
 VERSION AX002804.1 GI:9885132
 KEYWORDS
 SOURCE
 ORGANISM
 porcine endogenous retrovirus.
 Porcine endogenous retrovirus
 Viruses; Retroviral viruses; Retroviridae; Mammalian type C
 retroviruses; 1-Mammalian type C virus group.
 Stoye, J.P. and Weiss, R.A.
 Detection of retroviral subtypes based upon envelope specific
 sequences
 JOURNAL Patent: WO 9853104-A 3 26-NOV-1998;

MEDICAL RES COUNCIL (GB); STOYE JONATHAN PAUL (GB)
 Location/Qualifiers
 1.3482
 /organism="porcine endogenous retrovirus"
 /db_xref="taxon:61673"
 BASE COUNT 927 a 854 c 867 g 834 t
 ORIGIN
 Query Match 59.0%; Score 1743; DB 6; Length 3482;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 2913; Conservative 0; Mismatches 21; Indels 1; Gaps 1;
 QY 1 tggcttgggtttagaagaccccttgagcaatttgggctgaacccctatgattgctct 60
 DB 548 TGGCTTTTAGGGTTAGGAACCCCTGAGCAATTTGGCGTGAACCCCTATGAATTCCT 607
 QY 61 acggggaaccccccgttggtagaattgtctctgataagttgtaigtgctgtt 120
 DB 608 ACGGGGAGCCCCCGCTTGGTGAATTCCTTGTACATAGTCTGATGCTGCTTT 667
 QY 121 cccagccttctctctagctcaagcgctcagtggttggagcaacgagctgtgaaac 180
 DB 668 CCCAGCTCTGTTCTTAGGCTCAAGCGGCTGAGTGGGTGAGGCAAGCGTGAAGC 727
 QY 181 agctccggagagcttactcagagagagagagcttgcagttccacatcgcttcaagltg 240
 DB 728 AACTCCGGAGGCTCTACTCAGAGAGAGACTTTCAGAGTTCCACATGCTTCCAAAGTGG 787
 QY 241 gagattcagctatgtttagacgcccagctgcaagaaacctgagactggtggaaggagac 300
 DB 788 GAGATTCAAGTGTATGTTAGAGCCCACTGCAAGAACTTGAAGTCTGAGTGAAGGGCC 847
 QY 301 ctatctcgttactctttagacacacacacagcgctgtgaagtgcgaagatctcaccctgga 360
 DB 848 CTATCTCTGACTTGTGACCAACACAGCGCTGTGAAGTGAAGGAATTCACCTGGA 907
 QY 361 tccatgcaatcccaagcttgaagcttgagccactccagctcgggtgtgaaagcggaaga 420
 DB 908 TCCATGCAATCCCACTTAAAGCTTGGCGCACCTCCGACTCGGCTGGAAGGCCAAGAA 967
 QY 421 ctggaatcccttaagcttgcctcactcagctcgtgtgttcttacttcaacaataactccc 480
 DB 968 CTGAATAATCCCTTAAGCTTGCTCCATGCGCTGCTGCTTCTACTTACATAAATACCTCC 1027
 QY 481 caggccagtagtaaacgcttatagaagctcgaaaccccatgaacttaccctacc 540
 DB 1028 CAGGCGAGTAAACGCTTATAGACAGCTCGAACCCCATGAGACTTATCCCTTACC 1087
 QY 541 tggctgattatgacccctgatacgggtgtcactgttaataagacatcgaggttgttgcct 600
 DB 1088 TGGCTGATTATGACCTGTGATACGGGTGTCACTGTAAATAGCACTGAGGTGTGCTCT 1147
 QY 601 agaggaacctgtgtgacctgaactgcatcttgcctccgattgtatlaaccgcgtgtlaaa 660
 DB 1148 AGAGGCACTGTGTGGCTCTGAATGCAATTCGCTCCGATTGATTAAACCCGCTGTAA 1207
 QY 661 agcacacctccaaacacagcgttagttataggttctatctgtcgcagggcacagaaga 720
 DB 1208 AGCAACCTCCCAACCTGAGTCCGTAGTATGAGGTCTTATTTGCGCCACAGCAAGAA 1267
 QY 721 gagaataactgttggggttcttgggaatccctgttagagagatggagctgtcctacc 780
 DB 1268 GAGAAATACGTGTGGGTTGTGGGAAATCTTGTGTAGAGATGAGAGCTGCTACACTCC 1327
 QY 781 aacagatggaactgtgaatgtgcccagatcctctccagagacgggtlaaatctccttctc 840
 DB 1328 AACGATGGAACCTGGAAGAGGCGAGTCTCTCCAGACGCGGTAATTCCTTGTGTC 1387
 QY 841 aattcggcgccgggcaagtaacaaatgataactataaagataagagctgtctccca 900
 DB 1388 AATTCCGGCCCGGCAAGTACAAAGTATGAATAAAGATTAAGATTAAGAGCTGCTCCCA 1447

Qy	901	tcagacttagatlatctaagaatglttcaactcyaaaagaaacacaggaataatltcaa	960
Db	1448	TCAGACTTAGATTATCTAAAGTAAAGTTTCCTCAAAAMAGAAACAGGAATAATTTCA	1507
Qy	961	aagttagataatggtatagatgcctggagaaatglttttaataatgctggggagcagag	1020
Db	1508	AAGTGAATAATGATGATGAGCTGGGGAAATAGTTTTTATATAATATGCGGGGAGAGAGG	1567
Qy	1021	lccaacttaaccattgcgcttiagaatagaacggygagacgaacccccgtlygcaatggga	1080
Db	1568	TCGACTTTAACCTATGCGCTTGAGTATGAGACGGGGGACAAACACCCCTGTGGCAGTGGGA	1627
Qy	1081	cccgataaagatctgctctaaagaagggccccggcccccttgagccacgcgaataacttgcg	1140
Db	1628	CCCGATAAATACTGCTCTAAACAGGGGGCCCCGGCCCTGGAGGCCACCGCATMACTTGCGC	1687
Qy	1141	gtgcaccaacttaaccctcgcgtgcygcctiyacaatacacagccgcctagcaacagtlaccact	1200
Db	1688	GTGCGCCCAATTAACCTCGCTGGCGCTGACATTAACACAGCGCGCTTAGCAACGCTTACCT	1747
Qy	1201	ggattgatttctctaccaacgcgctagaacctccccaggtgttctctgttaagaacagacag	1260
Db	1748	GGATTGATTCCTTACCAACACGCGCTGAAATCTCCAGGTGTTCCTGTTAAGACAGACAG	1807
Qy	1261	agactcttcagttcatccaggagacgtltccaagacalcaactccaacgcgcctgaltgc	1320
Db	1808	AGACTCTTCAGTGCATCCAGGGAGGTTCCAAAGCATCACTCCACCGACCTGTAGTGC	1867
Qy	1321	actctctctgttggtcgtttgtctatccctcaaggcctccttlatitgagggatggtcttaa	1380
Db	1868	ACTCTCTCTGTGGCGTTTGCTATCTCAGGGCGCTCCTTATATAGAGGGATGGCTAA	1927
Qy	1381	gaagaagaatctcaatgtygaccaagaagcaagaataatcatatgcatggyggtlccggaat	1440
Db	1928	GAGGAAATTCATGTGACCAAGACATTAATAATCATGTACATGGGGGTCCCGAAAT	1987
Qy	1441	aagcttaacctactagaagtttcggygaagggagacatgcatagaaaaagcttccccatcc	1500
Db	1988	AAGCTTAACCCCTACATGTAAATTTCCGGAGGGGACATCATAGGAAAGCTCCCCATCC	2047
Qy	1501	caccaaacaccttggctatagactgtggttlatitgycacgagcctcagaanaatcgaatlla	1560
Db	2048	CACCAACACTTCTTCTTATGACTGTGTTTATGACAGCGCTCAGAAATACATATTTA	2107
Qy	1561	gtacctgtgltlaaacaggtggtggygacatgcaactggtglttaeccctgtgttccacc	1620
Db	2108	GTACTGTGTTATACAGGTGGTGGCATGACATACGAGGTATACCCCTGTGTTTCCACC	2167
Qy	1621	tcagcttccaacccatccaagaaglttgggtgtcatggtccaaatgcgtcccccaggttacc	1680
Db	2168	TCAGCTTTCACCAATCCAAAGATTTGTGTGTATGTGTCCAAATGCTCCCCGAGTGTAC	2227
Qy	1681	taacctctcggaagatggttctctatgaaatlatgactcgtgtataccgacccaanaaga	1740
Db	2228	TACCAATCCGAGGAAGTGTCTTGATGATATGATATCGGTATTAACCGACCAANAAGA	2287
Qy	1741	gaaccctgatccctaccctcagctgtatggtctgggtttgagagggcggtttggtgttga	1800
Db	2288	GAAACCGTATCCCTTACCTTAGCTGATATCTCGGATTAAGGAACGCGCGTGTGGCTAGGA	2347
Qy	1801	acaggaagacgtgcgcctbatcacagagccacagcagctlagagaagaagacttgttagcta	1860
Db	2348	ACAGGACACGCTGCCCTGATCCAGGACACACAGCAGCTTAGAAGAGAAAGACTTGTAGCTA	2407
Qy	1861	catgagccatgacagaagaatctccgagccttaagaggtctgtttagaacttagaag	1920
Db	2408	CATGCGGCANTACAGGAAGATCTCCGAGCCTTAGAGAGATGTGTTAGCACTTAAGACAG	2467
Qy	1921	ttcccttaactcttcttgcgaagtgttctacaagaacggaagggatlagatctgctt	1980
Db	2468	TCCCTTACTCTTTGTCTGAAAGTGGTCTACAGAACGGAGGGGATTAATCTGCTGTTT	2527
Qy	1981	ctaaagaaagtgggtaltatgltgcagccttaaaagaagaaltgtgctctatgtagataac	2040

Viruses: Retroviral viruses; Retroviridae; Mammalian type C
 retroviruses: 1-Mammalian type C virus group.
 REFERENCE
 1 (bases 1 to 3482)
 Le Tissier, P., Stoye, J.P., Takeuchi, Y., Patience, C. and Weiss, R.A.
 TITLE
 Two sets of human-tropic pig retroviruses
 JOURNAL
 Nature 389 (6652), 681-682 (1997)
 MEDLINE
 97478526
 REFERENCE
 2 (bases 1 to 3482)
 Stoye, J.P.
 AUTHORS
 TITLE
 Direct Submission
 JOURNAL
 Submitted (02-APR-1997) J.P. Stoye, National Institute for Medical
 Research, Virology, The Ridgeway Mill Hill, London, NW7 1AA, UK
 REMARK
 revised by author 16-SEP-97
 FEATURES
 location/Qualifiers
 1..3482
 /organism="porcine endogenous retrovirus"
 /specific_host="Sus scrofa"
 /db_xref="taxon:61673"
 /note="PERV-B"
 911..2884
 /gene="env"
 911..2884
 /gene="env"
 /codon_start=1
 /protein_id="CAA72928.1"
 /db_xref="GI:2576325"
 /db_xref="SPRMBL:O41173"
 /translation="MPTLSMRHLPTRGEPKRLRIPISFASIAFWLTLITTPQASSK
 RLIDSNRHLPLJLWLIIDPTGVTVSTKGVAGRWMPDLHLCLILNPVASTP
 PNLVRSYGCPCPEKEKYCGSGESGSRSCVTSNDGKMPISLDIRKYSFVN
 SGPGRYKWKLEKDKSCSPSDLYLKISFTEKQENIOKMTSMGIVRYKKYCGA
 GSTILRLRIETGPPVAPVADKVLASIOGAPALPNNLPVPLTSTRIPDTPPSN
 GTTILPNTSPSPVYKVGORLPSILOGAPOLINSTDPATSSCWLCSGPPY
 EEMAEKGNVYKRNOCNCTMSRKLTLTEAGSGCTIGAPSPHOLCYSTVYEQ
 ASENQYLVPEVYKRWACNTGLTPCVSTVFNQSKDPCVVOIVPVSVYHPEVLEDEY
 DYRNPKEPVSILAVMLGLTAVGVTALITGPOLKEGLGELHAAMEDLR
 ALEESVNLSELSISLVVONRGLDLLEREGALCAALKKEGCFVVDHSGAIRD
 MSKLRELERREREREDOGWFEFENSPMTTLTSLGTPLVYLLLLVYGCCLIN
 RFVAFVRVSAVOIMVLRQYOGILSGENDL"

BASE COUNT 927 a 854 c 867 g 834 t
 ORIGIN

Query Match 59.0%; Score 1743; DB 14; Length 3482;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 2913; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

QY 1 tgccttttaggttaggaacaccccttggaagcttgagctgacccccctataatgctct 60
 DB 548 tgccttttaggttaggaacaccccttggaagcttgagctgacccccctataatgctct 607
 QY 61 acgagggaacccccctgtgtagaattgctctgtacatagtgctgctgctt 120
 DB 608 acgagggaacccccctgtgtagaattgctctgtacatagtgctgctgctt 667
 QY 121 cccagccttgcctctcagctcgaagcgtcgaagtgagtgagcaagcgtggaagc 180
 DB 668 cccagccttgcctctcagctcgaagcgtcgaagtgagtgagtgagcaagcgtggaagc 727
 QY 181 agctccgggaagcctcagcgaaggaagagcttgcaagctccacatgcttccaagt 240
 DB 728 agctccgggaagcctcagcgaaggaagagcttgcaagctccacatgcttccaagt 787
 QY 241 gagatcagctatgtagaagcagcgtgcaagaaactcgaagctcogtgggaagac 300
 DB 788 gagatcagctatgtagaagcagcgtgcaagaaactcgaagctcogtgggaagac 847
 QY 301 ctatctcgtactctttagcaacacgaagcgtgaaagtgcaagaaactcgaagc 360
 DB 848 ctatctcgtactctttagcaacacgaagcgtgaaagtgcaagaaactcgaagc 907
 QY 361 tcaatgacccaggaagtgagcgcacactcccgagctcgagtgagagccgaaga 420
 DB 1988 tcaatgacccaggaagtgagcgcacactcccgagctcgagtgagagccgaaga 420

DB 908 tccatgcatccacggttaacgctggcgccacgctccgacatccgggtgagacccgaaga 967
 QY 421 ctgaagatcccttaagcttcgctccacacgctggtcttcttaacttaacaatactcc 480
 DB 968 ctgaagatcccttaagcttcgctccacacgctggtcttcttaacttaacaatactcc 1027
 QY 481 caggcagtaglaaacgctctatagacagctcgaacccccctataagcttaccctacc 540
 DB 1028 caggcagtaglaaacgctctatagacagctcgaacccccctataagcttaccctacc 1087
 QY 541 tggctgattatgacccctgatacagcgtgtaactgtaataagactcgaagtgctcct 600
 DB 1088 tggctgattatgacccctgatacagcgtgtaactgtaataagactcgaagtgctcct 1147
 QY 601 aagagcactggtggtcgtgaactgattctcgtccgattgataaacccgctgtaaa 660
 DB 1148 aagagcactggtggtcgtgaactgattctcgtccgattgataaacccgctgtaaa 1207
 QY 661 agacacccctcccaactagtcgtagtattggttctatctgctgcgaggaagaa 720
 DB 1208 agacacccctcccaactagtcgtagtattggttctatctgctgcgaggaagaa 1267
 QY 721 gggaaatactgtggggtcttgaggaaactctctgtaagagatggagctgcaactcc 780
 DB 1268 gggaaatactgtggggtcttgaggaaactctctgtaagagatggagctgcaactcc 1327
 QY 781 aacgataagcagtggaatgagcgaatctctccagagccgggtgaaatcttcttctc 840
 DB 1328 aacgataagcagtggaatgagcgaatctctccagagccgggtgaaatcttcttctc 1387
 QY 841 aatccgagccggggaagtagaataatgtagaactataaagaataagctgctccca 900
 DB 1388 aatccgagccggggaagtagaataatgtagaactataaagaataagctgctccca 1447
 QY 901 tcaagcttaggttcttcaagaataagttcaactgaaagaaagaaagaaatattcaa 960
 DB 1448 tcaagcttaggttcttcaagaataagttcaactgaaagaaagaaagaaatattcaa 1507
 QY 961 aagtgataatgtagttagcgtgggaatgatttttaataatgagcggggaagcagg 1020
 DB 1508 aagtgataatgtagttagcgtgggaatgatttttaataatgagcggggaagcagg 1567
 QY 1021 tcaacttaacacatgcttgaagtagaagcaggggaacacccctgagcaatggga 1080
 DB 1568 tcaacttaacacatgcttgaagtagaagcaggggaacacccctgagcaatggga 1627
 QY 1081 cccgataagtagctgctgtagaagcaggggccccggcccttgaagcagcaactgacg 1140
 DB 1628 cccgataagtagctgctgtagaagcaggggccccggcccttgaagcagcaactgacg 1687
 QY 1141 gtagcccaatlaacactgctgtagcagcagacatacaagcgcgctagcaacag 1200
 DB 1688 gtagcccaatlaacactgctgtagcagcagacatacaagcgcgctagcaacag 1747
 QY 1201 gtagttagtctctcaacaacagcctagaacactcccaagtgcttctgtaagcagca 1260
 DB 1748 gtagttagtctctcaacaacagcctagaacactcccaagtgcttctgtaagcagca 1807
 QY 1261 agactctcagctcagcagagagcttcccaagcagcaactcccaagcagcagctgac 1320
 DB 1808 agactctcagctcagcagagagcttcccaagcagcaactcccaagcagcagctgac 1867
 QY 1321 actctctctgtgcttgcctatccctcagagcctcctatataagagggatgagctaa 1380
 DB 1868 actctctctgtgcttgcctatccctcagagcctcctatataagagggatgagctaa 1927
 QY 1381 gaaagaaatlaaatgtagcgaagaaagagatagaatcaatgtagatggggtccgaaat 1440
 DB 1928 gaaagaaatlaaatgtagcgaagaaagagatagaatcaatgtagatggggtccgaaat 1987
 QY 1441 aagcttaacactcagtaaatctccgggaagggagacatgataagaaagctcccaatcc 1500
 DB 1988 aagcttaacactcagtaaatctccgggaagggagacatgataagaaagctcccaatcc 2047

CDS		2120..4726
		/gene="pol"
		/codon_start=1
		/protein_id="CA76582.1"
		/db_xref="GI:3005967"
		/db_xref="SPTREMBL:O73505"
		/translation="MGAQISFEDGKREVSNNKPTVTLTQLDDEYRUYSPQVKEDD IQSMLEDFPQMAEETAGMGLAKOVPOVITOLKASATPVSVROPLSRAREIMPHVQ RLIQGLILVAVOVSQWNTPLPVKPGSTNRYKQVODIRENKVODIHPTEPVLIS ALPPEKMTVDLKDAPFCLRLHPTSOPLFAFEMRDEPTGSGTOLMTPLRPOGKNS PTIFDALHDLNFRIOHPVITLOVYADKLILACATGDCCEGKALLLESDIGTR ASAKKAOICRREVTYGLSGRGORLTPTARKKTYVOIPAPTTAKOVEFSTGFCFR LMTFGPATLAALPLPTKEKESPMAPHEOKAFDIKRALISAPLALPDYTKFPTLY VDERKVAAGVLOTLGPMWRPVAATSKLDVAVSGMPCVKALIAVALIVKADKLIT LGNITVIAHALENTVROQPDMMTNAMTHVOSLILTERTEAPALNPAITLLE ETDEPYTHDCHOLLIEETGVKRLDIDPLTEGEVLTWPTDSSYVEGRMGAAYVOC THTIMASSLEPETSQAKELMALTOALRAEKSINITYDSRYAPATPHVGAIVYOR GLTTSAGREIKNKEETLSLEALHLPKRLATIHCHGHRKADLISRNOMADRYAKOR AONVNLPTIEMPRAPERROTYLLEDMOIKITIDFSETPECTYSTYKELPHKES LEYVOOIHNLTHGTRKLDQLVKISPYHVLRLPGVADSVVHCVPCOLVNNPRTIP GRLRSHPGAHMEVDTEVKPAKYGNNLYLVFVDFSGWEAVPTKETSTVAKKY WRKFFDLEYLK"
gene		5294..7267
		/gene="env"
		/codon_start=1
		/protein_id="CA76583.1"
		/db_xref="GI:3005988"
		/db_xref="SPTREMBL:O73506"
		/translation="MHPTLSMRHLPTRGGEPRRLRIPISPAISWFLTLTTPQASK RLIDSNPRLSLTWLITIDPDGYVSTGTGVAIPRGTWMELEHFCRLINPAKSTP PILVRSYGCPCPTEKERYCGSGESCTKRWSCVTSDSGDKWPIISLOHVKESFN SGPCKRYKWLTKDKSCSPSDLYKISFETKQEMINOKMINMSGIIFYXYGGA GSTLIRLLETGEPVAVGPDVLAEOGPALPEPHNLVPLDITSLRDIOPEN GTLPLINTPNSPSPGVPTGQRLFSLIQGAFAISTDPDASTSCMLDSSPPY EGMAKEGFENVYKEHKNQCTWGSNRKLLTFEVSQICIGAPASHOHLCTYVYVVO ASENOYLVPYMRMACNTGLPCVSSSVENOSDPCVMQIYVRVYHPEVYVLEO DYRNRPKREPVSLTAVMLGIGTAVGCGTALITGPOLKEGLDELHAAMTEDEIR ALEESVNLSESTISLSEVYLQNRGDLPLRLREGGLCALKECCCYVHSGAIRNS MSKLEHRELRERREHEDQNFEGMFPNSPMTWLTLSALGLPVLVLLLTLPSCILIN REVAFRERVAVOQIMVLRQYQGLISQGETDL"
CDS		5294..7267
		/gene="env"
		/codon_start=1
		/protein_id="CA76583.1"
		/db_xref="GI:3005988"
		/db_xref="SPTREMBL:O73506"
		/translation="MHPTLSMRHLPTRGGEPRRLRIPISPAISWFLTLTTPQASK RLIDSNPRLSLTWLITIDPDGYVSTGTGVAIPRGTWMELEHFCRLINPAKSTP PILVRSYGCPCPTEKERYCGSGESCTKRWSCVTSDSGDKWPIISLOHVKESFN SGPCKRYKWLTKDKSCSPSDLYKISFETKQEMINOKMINMSGIIFYXYGGA GSTLIRLLETGEPVAVGPDVLAEOGPALPEPHNLVPLDITSLRDIOPEN GTLPLINTPNSPSPGVPTGQRLFSLIQGAFAISTDPDASTSCMLDSSPPY EGMAKEGFENVYKEHKNQCTWGSNRKLLTFEVSQICIGAPASHOHLCTYVYVVO ASENOYLVPYMRMACNTGLPCVSSSVENOSDPCVMQIYVRVYHPEVYVLEO DYRNRPKREPVSLTAVMLGIGTAVGCGTALITGPOLKEGLDELHAAMTEDEIR ALEESVNLSESTISLSEVYLQNRGDLPLRLREGGLCALKECCCYVHSGAIRNS MSKLEHRELRERREHEDQNFEGMFPNSPMTWLTLSALGLPVLVLLLTLPSCILIN REVAFRERVAVOQIMVLRQYQGLISQGETDL"
BASE COUNT		2077 a 1971 c 2047 g 1713 t
ORIGIN		
Query Match		56.0% Score 1654; DB 14; Length 7808;
Best Local Similarity		99.2% Pred. No. 0;
Matches 2854; Conservative		0; Mismatches 24; Indels 0; Gaps 0;
QY	1	tgcttttaggttagaacaccccttgacagtttggttgaccccccattgaattgctct 60
DB	4931	tgcttttaggttagaacaccccttgacagtttggttgaccccccattgaattgctct 60
QY	61	acggggggaaccccccgcttgtagaattgctctctgacatagtgctgctctt 120
DB	4991	acggggggaaccccccgcttgtagaattgctctctgacatagtgctgctctt 120
QY	121	cccaagccttftctctgagctcaagcgctcgagtggtgagggcaacgagcgttgaa 180
DB	5051	cccaagccttftctctgagctcaagcgctcgagtggtgagggcaacgagcgttgaa 180
QY	181	agctccggagagcctactcaagagagagagcttgcaagtctccacatcgcttcaagt 240
DB	5111	agctccggagagcctactcaagagagagagcttgcaagtctccacatcgcttcaagt 240
QY	241	gagatcagctctatgttagagcgcaacgctgcaagaaacccctgagactcgttggaagg 300
DB	5171	gagatcagctctatgttagagcgcaacgctgcaagaaacccctgagactcgttggaagg 300
QY	301	cttactcgtacttttagacacacacacgctgtggaagtcggaagaaatccacctgga 360
DB	5231	cttactcgtactttttgacacacacacgctgtggaagtcggaagaaatccacctgga 360
QY	361	tcacatgcatcccaagcttaagctgagccacccctccgaactcgggggtgagagacggaaga 420
DB	5291	tcacatgcatcccaagcttaagctgagccacccctccgaactcgggggtgagagacggaaga 420
QY	421	ctggaatccccccttaagcttcgacctcgaactgagcttgcttacttaacataactccc 480
DB	5351	ctggaatccccccttaagcttcgacctcgaactgagcttgcttacttaacataactccc 480
QY	481	caggccagtagtaaacgccccttagaagcgtcgaaaccccatagacacttaaccctacc 540
DB	5411	caggccagtagtaaacgccccttagaagcgtcgaaaccccatagacacttaaccctacc 540
QY	541	tggtctgattttagaaccttgtagcgggtgctacgtgtaataagacactcgaagtgtgctcc 600
DB	5471	tggtctgattttagaaccttgtagcgggtgctacgtgtaataagacactcgaagtgtgctcc 600
QY	601	agaagcaacctggtgctgtagcgaactctctcgtccctcgaacttgattaaacccgctgtaa 660
DB	5531	agaagcaacctggtgctgtagcgaactctctcgtccctcgaacttgattaaacccgctgtaa 660
QY	661	agcaacacctcccaacctagtcogtagttaagggtctatgtctgctccagagacagagaa 720
DB	5591	agcaacacctcccaacctagtcogtagttaagggtctatgtctgctccagagacagagaa 720
QY	721	gagaataactgtggtggtctctgaggaactctctctgtagagaatgagctgtaacctcc 780
DB	5651	gagaataactgtggtggtctctgaggaactctctctgtagagaatgagctgtaacctcc 780
QY	781	aacgataagagactgtaaatgagcgaactctctccagagacgggttaaatctctctgtc 840
DB	5711	aacgataagagactgtaaatgagcgaactctctccagagacgggttaaatctctctgtc 840
QY	841	aattccggccggcggaagtaacaaatgataagaaactataaagataagagctgctcccca 900
DB	5771	aattccggccggcggaagtaacaaatgataagaaactataaagataagagctgctcccca 900
QY	901	tcagacttagatcatctaaagaataagttcactgtaaaagagaaacaggaatattcaca 960
DB	5831	tcagacttagatcatctaaagaataagttcactgtaaaagagaaacaggaatattcaca 960
QY	961	aagtgataataatgataatgagcgtgggaataagttttaaataatatgagcgggagagga 1020
DB	5891	aagtgataataatgataatgagcgtgggaataagttttaaataatatgagcgggagagga 1020
QY	1021	tcacacttaacatccgaccttagagatagagacgggagaaaccccttgtagcaattgga 1080
DB	5951	tcacacttaacatccgaccttagagatagagacgggagaaaccccttgtagcaattgga 1080
QY	1081	cccgataaagtaactgagctgtaaacagggcccccggcccttgtaggacacgcaataactgcg 1140
DB	6011	cccgataaagtaactgagctgtaaacagggcccccggcccttgtaggacacgcaataactgcg 1140
QY	1141	gtgcccacatctaactctgctgctgagctgacataacacagccgcttagcaacagtaacct 1200
DB	6071	gtgcccacatctaactctgctgctgagctgacataacacagccgcttagcaacagtaacct 1200
QY	1201	ggattgattccttacaacagccttagaagaccccccaggttctctgtrtaagaaagagag 1260
DB	6131	ggattgattccttacaacagccttagaagaccccccaggttctctgtrtaagaaagagag 1260
QY	1261	agacctctcagttctcatccagggagccttccaagcctcaactcacaacccgacccctgtg 1320
DB	6191	agacctctcagttctcatccagggagccttccaagcctcaactcacaacccgacccctgtg 1320
QY	1321	acttctctcttggtgcttggtgcttactcctgagggcctctctataatgagggagatgctaa 1380
DB	6251	acttctctcttggtgcttggtgcttactcctgagggcctctctataatgagggagatgctaa 1380
QY	1381	gaagaaataatcaatgtgacacaaagacatagaataatcaatgataatgagggagatgctaa 1440
DB	6311	gaagaaataatcaatgtgacacaaagacatagaataatcaatgataatgagggagatgctaa 1440

/db_xref="SPTREMBL:0901X5"
/translation="MGCTVTPPLSLTLDHWTBYRSAHNLSNOVKKGMOTFCASWPT
TFVGMSESTENSEILLAKAILFQTPGSHSDQEPYLLITWDLAEDPPWPKMLN
KPRKPGRIILAEKNHSAEKEVPSRIYPLEIEEPTMEPEPPVPAQAGVAG
PSAPGAPVVEGAGAGRSRGAATPRTDIALILPRITGPPMGOLPDIOWPSS
ADLYNMKTNHPEPSEDORLGLVESIMESHOTMTDQCOLLTLTTEREELILEA
RKVPAGDGRPOLONEIDMGPLTGPMDYNTAEORESLKITROLVAGLRASRIR
TNLAKEVWOGNEPVSFLEKMAEFRTIPDTSEKOKSAVALATGOSALDIR
KTLRLGELAEALRLDLYRAEKVYIRRETEERKORKEEREREERERORKEMLT
KILAAVEGSSREBRERFRKIRSGROSGNLNRPDLKQACVCKEGHARMLPK
KGNKGKVLTEEDKDRGSDPLPEPRVYLKVEQVPEVLPGLHSDLOPLK
KEKSMWGAATGROYWMTRTVDLAVGVTHSPFVTECPVPIAGHRLTKMSOI
SFGQREPVSNKPIITVLITLDDDEYRLSPQVCKDOOLQSLBEFPQNAETAMG
LAKQVPOVQLKASATPVSYROTISREKREPTWVHRLQGLVHVPWMTPL
LPKREPTNDYREVVDLREYNKRVODIHVPVPMYLSALPERWVYVLDLKDFE
CLRHLHSOPLFAFWMDPQGTGQTLTWRLPQGRKNSPTIDEALHDLNFRLOH
POVTLQYVDDLALAGATKODLEGTKALLLESLDGRASAKKQICREVEYLYS
LRGQRRLTEARKTYVOIPAPTTAKOVREPLGTACRMLMGPFTLAVIPLTKE
KGESNAPEHOKAPDAIKKALLSAPALALPDYTKPPTLYDEKGVARVLTOTLGM
RRPVYLSKRLDPYASGWPCKALNAVALVAKDADKLVGONTIVYARHLENIVRO
PDRWMTNMTYHQSLLTERVTFAPPALNATLLEPTEDEPVHDLHLEETG
VRKLDNDPLTGVLTWTDGSSVVEGRMAAAVVDGHTITWASLEGSAQAE
LMALTOALRLAEKGSINITYDSRYAFATVHGAIKOKGLTASREIKNEEILSL
LEALHLEKRLAIHCPGHOKAKDLISGNOMADRVAKOAAVNLPIETPKAPEPR
ROYTLEDMOIKRIDOESPEEGTCTSYGKETLHPKEGLVYQOITHTLHGTAKLO
QIVRTSPYHILRPGVADSVYKCVPCQLYNANPSIIPGKRLRGSHPAHEVETE
VFAKTGNKILVFEVDFSGWVAEYPTKETSTVNAKILLEETFPREGPKVIGSNG
PAFAVOVSQGLALIGIDMKLHCAYRPOSSGOVEREARNADYLETILKLTETENMA
LLPFLVFRVNTPGOGLPYELLGPPPLVETASVHADYVLSQLEPRLKALEM
KORAMKOLREAYSGEDLOVRFQYDGSVYVRHAGNLETMSKPPYVLLTPTTAV
KVGESITWHTSHVKLAPPDGSMRAKEMENPKLHLVLPVSNKNSPGOLHPTTSM
RHLPTGSGPKRRLPLSPASIAMPLTITITTPQASKRLIDSNRPRLSLTWLITDP
DTGVTYNSTRGAPRGTMPELHFCULINPAKSTPPMLVRSYGVCCPGKEKEXYK
GDSGESFCRMSCVTNDGDKWPILODRKVSFNQSPGKRYWKLCDDKSCPSD
LDILKISFTEKPOENIOKWINMSWGIYFYXGAGSTLILRLITETGPVAVG
PDVLAEGPPLPEPHNLPEPLTSLRDIOSPGNGTGLIPNTPNSGVAPKT
GORLFLSIOGAFOAINTSDPDATSSCMVLPYDEAMKATVYKTRKRNOCOTW
GSRNKLITLREVSCKGICIKAPPSHOLCTSYVYEQASBNOLVYKTRKRNOCOTW
TPCVSTSVENQSCDFCVVOYIVRYVTHEVEVLDYDRIYNPKREPVSLMACNGL
LGTAVGVTGTAAALITGPQOLEKGLGELHAAEMEDIRALEESNLESLTSEVLL
QNRGRDLFLREGGLCAALKECCFCYVHSGAIRMSKLESLERLERERREADDGW
FEGMFRSPMTLLSALTGPLVLLVLLLVGLVCLINRVAVFREVSAVOIMVLROQ
YOSLISQSGERDUL
2729. 6313
mat_peptide
/gene="pol"
/function="retroviral protease and polymerase"
/product="pol protein"
2729. 6313
gene
/gene="pol"
6189. 8162
mat_peptide
/gene="env"
/function="retroviral envelope"
/product="Env protein"
6189. 8162
gene
/gene="env"
8212. 8918
LTR
8811. 8816
polyA_signal
2337 a 2245 c 2334 g 2012 t
BASE COUNT
ORIGIN

Query Match 55.3%; Score 1636; DB 14; Length 8918;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2736; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 tgccttttaggttagaacacccctggaacagtttgagcctgacccctatgaattcctc 60
|||||
Db 5826 Tgccttttaggttagaacacccctggaacagtttgagcctgacccctatgaattcctc 5885
QY 61 acgagggagaccccccgttgtagaattcctctgacacagtagtgcgtatgctcctt 120
|||||
Db 5886 ACgagggagaccccccgttgtagaattcctctgacacagtagtgcgtatgctcctt 5945

QY 121 cccagccttctctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 180
|||||
Db 5946 CCCAGCCT 6005
QY 181 agctccggagagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 240
|||||
Db 6006 AGCTCCGGAGGCGCT 6065
QY 241 gagatcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 300
|||||
Db 6066 GAGATTCAGTATGTTAGACGCGACCGTGCAGAAACCTCGAGACTCGGTGGAAGGGCC 6125
QY 301 ctatctcgtactcttctgaccacacacacacacacacacacacacacacacacacacac 360
|||||
Db 6126 CTATCTCTGACTCTTTGAC 6185
QY 361 tccatcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 420
|||||
Db 6186 TCCATACATCCACGTTAAGCTGGCCGCCCTCCGACTCGGGGTGGAGCGCAAAAGA 6245
QY 421 ctgagatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 480
|||||
Db 6246 CTGAGATCCCTTAACTTTCGCTTCATCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 6305
QY 481 cagggcagtagtaaacgacctatagacagctcgaaccccaatagacacctatccctacc 540
|||||
Db 6306 CAGGCGAGTAGTAAGACGCTTATAGACACCTGGAACCCCATATAGACCTTATCCCTTAC 6365
QY 541 tgcctatattgacacctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 600
|||||
Db 6366 TGCTATATTATGACCTCTATACGGGTGTCATCTGTAATAGCACTCGAGTCTCTCTCT 6425
QY 601 agagacacctgctgagcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 660
|||||
Db 6426 AGAGGACACTCGTGGGCTGACACTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6485
QY 661 agcacacccccaactcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 720
|||||
Db 6486 AGCACACCTCCCAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6545
QY 721 gaaataactcctgagaggtcctcctcctcctcctcctcctcctcctcctcctcctcctc 780
|||||
Db 6546 GAAATAACTCTGAGGAGTCTCTGAGGATCTCTCTGAGGATGAGAGCTGCTCCTCTCC 6605
QY 781 aacgagtagagctggaatcctcctcctcctcctcctcctcctcctcctcctcctcctc 840
|||||
Db 6606 AACGATGAGACTGGAATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6665
QY 841 aattccgagccgagcaatgataaataatgataaataatgataaataatgataaataatg 900
|||||
Db 6666 AATTCCGAGCCGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6725
QY 901 tcagactgattatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 960
|||||
Db 6726 TCAGACTTATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 6785
QY 961 aagtgataaactgatacctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1020
|||||
Db 6786 AAGTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 6845
QY 1021 tccacttaacacatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1080
|||||
Db 6846 TCCACTTAAACATTCGCTTATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 6905
QY 1081 cccgataaagatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1140
|||||
Db 6906 CCCGATTAAGTACTGCTGTAACAGGAGGCGCCGCTGAGGACCAACCCGATTAATCTT 6965
QY 1141 gtcgcccaataactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1200
|||||
Db 6966 GTGCCCAATTAACTCTGCTGAGGCTGACATTAACACAGTGCCTTACGAAAGGATGCT 7025
QY 1201 ggaatgattcctacacacagcctcctcctcctcctcctcctcctcctcctcctcctc 1260

Db	6769	TCGCCTTAAACCATTTGGCCTT	TAGATATAGAGACGGGGACAGAAACCCCTGTGGCAGTGGGA	6828
Qy	1081	cccgataaagctactgctcgtctaaacgaagggcccccgcctcgagccacgcgcataactctgcg	1140	
Db	6829	CCGGATTAAGTACTGGCTGAAACAGGGGGCCCCCGGCCCTGGAGCCACCGCATAACTTTCGCG	6888	
Qy	1141	gtgcgccaatctaacctcgcctgcgcgcgcctgcatacaacagccgccttagcaaacagtaaccact	1200	
Db	6889	GTGGCCCAATTAACTTCGCTCCGCGCTGACATTAACACAGCCGCCCTAGACANAGTACACT	6948	
Qy	1201	ggatgtatctccctaacacacgcctagaanaactccccaagtgctcctgtttaagaacagacag	1260	
Db	6949	GGATTCGATTCCTACCAACACGCGCTAGAAACTCCCAAGGTGTCTGTTTAAACACGGAGACG	7008	
Db	7009	AGACTCTTCAGTCTCATCCAGGGAGCTTCCAAAGCATCAATCCACCGACCTCGATGCC	7068	
Qy	1321	actctctctgttgcgtcttgatcctccctcgaagccctcttaltagaagggatgtgtcaaa	1380	
Db	7069	ACTTCCTGTGTGGCTTGTGTCTATCTTCAGGCGCCCTTATTATGAGGGGATGGCTAAA	7128	
Qy	1381	gaaagaaatctcaatgtgacccaagagccataaatacaatgacatggaggtctcccgaaat	1440	
Db	7129	GAAGGAAATTCATATGAGCAAAAGGCGATTAATAATTCATGTACTGGGGGTCCCGGAAT	7188	
Qy	1441	aagcttaccctcaactgaagttccgcgaagggagacatgataagaaagctccccaatc	1500	
Db	7189	AAGTTTACCCCTCACTGAAAGTTTCCGGGAAGGGGACATGCATATGAAAAAGTCCCCCATCC	7248	
Qy	1501	caaccaacactttgctatagtaactgtgtttatagtcagagcctcagaanaatcaglatatta	1560	
Db	7249	CACCAACACCTTGTGCATATGACTGTGGTTTATGACACAGCGCTCGAAAAATCAGATTTA	7308	
Qy	1561	gtacctgttataacagatgtgtgagcatgaatactcgtgttlaacccctcgtgttccaac	1620	
Db	7309	GTACTCTGTTATTAACAGTGCTGGGGATGCAATACGCGTTAAACCCCTGTGTTCACCC	7368	
Qy	1621	tcagttcttcaaccaatcccaagaatctgtgtgtcaatgtgtccaaatcgtcccccagtgta	1680	
Db	7369	TCACCTCTTCAACCATCCAAAGATTTCTGTGCATGTGTCGCAATGTGTCGCCGAGTGTAC	7428	
Qy	1681	taccatcttaaggaagatgtgtccttgaatgaatagtaatcgtatcgttalaacccgcacaaaga	1740	
Db	7429	TACCATCTTGAGAGATGGTGCTCTTGATGATGATGACTATCGGTATTAACCGACCAAAAAGA	7488	
Qy	1741	gaaccgctatcccttaccctagctgtgaatgtctcgaaatlaggaacgcgcgtgtgcgtaga	1800	
Db	7489	GAACCCGATTCCTTACCCTAGCTAGCTGAATGCTGGAATGTGGAGACGCGCTGTGGCTAGGA	7548	
Qy	1801	acagaggacagctgtgcctcgtatccaagacccaacgcagctagaagaaagagctgtgtagta	1860	
Db	7549	ACAGGAGACGTCGCCCTGATTCACAGACCAACACACTGTAGAAAAAGACTTGTGTAGCTA	7608	
Qy	1861	catgcgcacatgaacagaagatctccgaagccttlaaagaagctcgtttagaacacttagaagag	1920	
Db	7609	CATGGGCGCATGACAGAAAGATCTCCGAGCCTTAGACAGACTGTTAGCAACCTAGACAG	7668	
Qy	1921	tcacctactcttgtctgaagtgtgtctacagaaacccgaagggagtagatctcgtcttt	1980	
Db	7669	TCCCTGACTTTCTTGTCTGTAAGTGTGTTCTAACAAACCGGAGGGGATTAATCTGCTGTTT	7728	
Qy	1981	cttaagaaaggttggtgtatcttgcaagccttlaaagaagaatgtgtcctcatgtatagatcac	2040	
Db	7729	CTTAAGAAAGGTGGGTTATGTGCAAGCCTTAAAGAAAGATGTTGCTTCTATGTAAATCAC	7788	
Qy	2041	tcaagagcatcagaagactccaatgaacaaacttagaanaaaaglttagaagagcgtcogaag	2100	
Db	7789	TCAGAGGCAATCAACAGACTCCATATGACCAACTTAGAAAAAGTTATAGACAGCGCTGGAAG	7848	
Qy	2101	gaagaagagcttgacccggaggttggtttgaagagtggttcaaaaggtcctcttgagtgacc	2160	
Db	7849	GAATAAGAGCTGTACACAGGGGTGGTTTGAAGATGTGTCAACAGGCTCTCTTGATGTAC	7908	

QY	2161	accctgcttctgctcttgacgagggcccttaatagtcctgcctccgttaactaacagttg99	2220
Db	7909	ACCCTGCTTTCCTGCTCTTACGAGGGCCCTTAATAGTCCTGCTCTGTTACTTACAGTTGGG	7968
QY	2221	ccttgcttaatlaaagaattglttgcccttgtlagaagaagagtgaatgacgtccagatc	2280
Db	7969	CCCTGCTTAATTAATAGGTTGTTGCTTTGTTAGACAACGAGTGAGTGAGTCACATC	8028
QY	2281	atggtacttaggcaacagttacaaagcctcttgacgaagagaacgtgaccttaagcct	2340
Db	8029	ATGGACTTAGGCAACAGTACCAAGGCTTCTGAGCCACAGAGAACTGAGCCTTAGCCT	8088
QY	2341	tcccgacttctaagaattagaactatlaaacaagaacaagaatgtygggaalgaagaagatga	2400
Db	8089	TCCCGACTTCTAAGATTAGATTAGACTATTAAACAAGCAAGAAGTGGGATGAAGAAGATGAAAA	8148
QY	2401	tgcacaactaacccctcccagaaccccagagaagtataaaaaacttlaaaltgccccgaatt	2460
Db	8149	TGCACAACCTTAACCTCCCGAAGACCCKGGAAGTTAATAAAAAACCTCTAATAATGCCCCGAAATT	8208
QY	2461	acagaacccctgctgctgctgcacgtlaaataagtlagaagtgatcaacttctatigtltccagggc	2520
Db	8209	CCAGAACCCTGCTGCTGCTGCACGTAATAATAGTGAAGAACGTCAACCTGCTATTGTTCCAAAGC	8268
QY	2521	ctgctatcctgagccctaaatagaataacaggaatgatgttgcttaatgccttatctg9att	2580
Db	8269	CTGCTATCCTGGCCCTTAAGTAAGTAATACACGGAAGTAATGATGTGACTAATGCTTAATCTGGGATT	8328
QY	2581	ctgtlaaanaactgacgtgcacacatagaagaagattgatatacatattgaagcccttaagtacc	2640
Db	8329	CTGTAAANAACCTGACGTGCACACATAGAGAAGAAATTGATTATACATTGACAGCCCTAGTGACCTTA	8388
QY	2641	tct 2643	
Db	8389	TCT 8391	

AF014162	1971 bp	mRNA	VRL	01-JUN-1998
LOCUS	AF014162	1971 bp	mRNA	VRL
DEFINITION	Porcine endogenous retrovirus envelope protein (env) mRNA, complete cds.			
ACCESSION	AF014162			
VERSION	AF014162.1	GI:3169735		
KEYWORDS	.			
SOURCE	porcine endogenous retrovirus.			
ORGANISM	porcine endogenous retrovirus			
	viruses, retroid viruses; Retroviridae; Mammalian type C retroviruses; I-Mammalian type C virus group.			
REFERENCE	1 (bases 1 to 1971)			
AUTHORS	Haworth,C., Galbraith,D.N., Lees,G.M. and Smith,K.T.			
TITLE	Porcine endogenous retrovirus (POEV) Env sequence			
JOURNAL	Unpublished			
AUTHORS	2 (bases 1 to 1971)			
TITLE	Haworth,C., Galbraith,D.N., Lees,G.M. and Smith,K.T.			
REFERENCE	Direct Submission			
TITLE	Submitted (16-JUL-1997) Q-One Biotech Ltd., Todd Campus, Acre Rd., Glasgow G20 0XA, UK			
JOURNAL	Location/Qualifiers			
FEATURES	1..1971			
source	/organism="porcine endogenous retrovirus"			
	/db_xref="taxon:61673"			
	/cell_line="derived from cell line PK15"			
	/note="POEV"			
	1..1971			
gene	/gene="env"			
	1..1971			
	/gene="env"			
	/codon_start=1			
	/product="envelope protein"			
	/protein_id="A017948.1"			
CDS	/db_xref="GI:3169736"			

/translaction="MHPTLSRHLPTRGEPKRLRPLSLFASIASLAMELTTLITPQASSK
RLIDSSNHPRLSLTJLWLIIDPDGTVNSTRGVAAPRQTNWDEHFLRLINPVAKSTP
PLNVRASYGFCPCPEKEKYGSGSGSESRWASVTNDGDMKPIELJQRYAFSEPN
SGPKYKMMKLYKDKSCSPDLIDYKLSIFERKTKSKYDKWELGNSPLITGGAG
STFLIRIENGTEPRPAMGPDYLAEROPALPEPNILPVOLITSLRPIITPPNS
TTGLIPTNPNSRGVYKQGLFSLIQARFOALNIDTDATSSCMLCISGPYTE
GMAKERFNVTKERNOCITWGRNKLTLTVESGCTIGAPSHOHLCSTVYEDQ
SENYLVPGYNRMWACNTGLTPEVSVFNQSDPCVMQIVPRVYHHPVEVLDEDA
YRNRPKEVPSLTLAVMLGILTAIVGLTALINGPOOLEKGLGLHAAEEDRA
LKEVSNNLESLTSLSEVLONRGLDLFLREGGLCAALKEGCEPVDHSGAIRDMS
NKLKLEERRREREADOGFEGFNPSNPMPTLISLTPVLVLLILLVGLCLNR
FVAFPRERVSANOIMVROOYOGILSGEETDL"

BASE COUNT 530 a 490 c 482 g 469 t
ORIGIN

Query Match 50.7%; Score 1499; DB 14; Length 1971;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

QY 383 ggccgacacccctccgactcggggtgagagccgaaagactgagatcccttaagcttcg 442
DB 20 GGGCCACCTCCGACCTCGGGGTGAGAGCGGAAAGACTAGATCCCTTAAGCTTCG 79
QY 443 cctccatcgctggttcttacttaacaatactcccgagccagtaagtaaacgctta 502
DB 80 CCTCCATCGCCTGGTTCCTACTCTAACATAAATCCCGAGCCAGTAAAGCCCTTA 139
QY 503 tagacagctgaaccccatagaccttatccctactgctgataattatgacctata 562
DB 140 TAGACAGCTGACCCCATAGACCTTATCCCTACCTGGCTGATTAAGCCCTGATA 199
QY 563 cgggtgtcactgtaaatagacactgaggtgttgccttagagagcactgtgtgacctaac 622
DB 200 CGGGTGTACACTGTAATAGACTCGAGGTGTGCTCTAGAGGACCTGGTGGGCTGAAAC 259
QY 623 tgcatttcgctccgattgattaccccgctgttaaaagcaacccctccaaactagcttc 682
DB 260 TGCAATTCCTCGCTCCGATGATTAACCCCGCTGTAAACACACCTCCCAACCTGATGCC 319
QY 683 gtaattatggttctattctgctgcagcagacagagaaagaaatactgtggtggttcgt 742
DB 320 GTATTATGGGTTTCTATTGCTGCCACAGCAGAGAAAGAAATACTGTGGGGGTTCTG 379
QY 743 gggaaatcccttctgtagagatgagatgagctgcacccctcaacgagtgagaaatgagc 802
DB 380 GGGAAATCCTTCTGTAGAGATGAGAGCTGCTCACTCAACGATVGGAGTGGAAATGGC 439
QY 803 cgaatctctccagagacgggtaaaatcttccttcttcaatcccgccgggcaagtata 862
DB 440 CGATCTCTCTCCAGAGCGGGGTAAATTCCTTTGTCAATTCGGGGCGGGCAAGTACA 499
QY 863 aaatgagtaaaactataaaagataagagctgctcccatcagacttagatctataaga 922
DB 500 AAATGATGAACATATATAAGATTAAGAGCTGCTCCCATAGACTTAGATTATCTAAAGA 559
QY 923 taagttcaactgaaagaaagaaataatcaaaagtggataaagtatgagct 982
DB 560 TAAATTTCACTG--AAAGGAAACAGGAAATATTCAAAAGTGAATGAATGATGAGCT 617
QY 983 ggggaataagtttttaataatagcggggagcagaggtccacttaacatccgctta 1042
DB 618 GGGGAATAGTTTTT-ATTATATGGCGGGGAGCAGGCTTCACTTTAAACCATTCGGCTTA 676
QY 1043 ggaatagagcgggagacgaacccctgtgcaatggagaccggaataagtaactggtcgaac 1102
DB 677 GGAATAGAGAGCGGGAGACGAAACCCCTGTGCAATGGGACCCGATTAAGTACTGGCTGAAC 736
QY 1103 aagggcccccggcctgagagcagcaataactgctggtgcccgaataaactgctgc 1162
DB 737 AGGGGCCCCCGCCCTGAGACCCGCAATTAATTCGGGGGCCCAATTAACCTCGCTGC 796
QY 1163 ggcctgaacataacagccgcttagcaacagtaacactgattgattctcctcaacaagc 1222

DB 797 GGCTGACATTAACAACAGCCGCTAGCAACAGTACACTGATGATTGATCCAGCAACAGC 856
QY 1223 ctgaaactccccaagtgttccgtttaaagacagagactcttcagttcaatccag 1282
DB 857 CTGAAACCTCCCAAGGTGCTCGTTAAGACAGAGACAGACCTTCAATCTCATCAG 916
QY 1283 gaacttccaagacatcaactcaacgcgaacctgtgcaacttctctgttgactgtgc 1342
DB 917 GACCTTCCAGCATCACTCACTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 976
QY 1343 tatccatagggcctcttattatagagggatggctaaagaaagaaatcaatgtgaca 1402
DB 977 TATCTCAGAGGCTCTCTTATTAGAGGGATGCTAAAGAAAGAAATCAATGTGACCA 1036
QY 1403 aagagatgaataatcaatgatactgaggggtcccgaaataagcttaaccccaagatt 1462
DB 1037 AAGAGCATGAATAATCAATGATGATGAGGGGTCCCGAAATTAAGCTTACCTCACTGAAGTTT 1096
QY 1463 ccgggaagggagacatgataagaaagctcccatcccaacaacaccttgcttaagta 1522
DB 1097 CCGGGAAGGGGACATGATGAGAAAGCTCCCATCCACCAACACCTTCTTATGTA 1156
QY 1523 ctgtgttataagcagcgcctcagaataatcagtaattagtaacctgtgtataaaggtgt 1582
DB 1157 CTGTGTTATAGAGCAGGCTCGAATAATCAGATTATGATACCTGTGTTATACAGGTGT 1216
QY 1583 gggcagcaataactggttlaacccctgtgttccactcagttctcaaccaatccaaag 1642
DB 1217 GGGCATGCAATATCTGGGTTAACCCCTGTGTTCACCTCACTTCAACCAATCCAAAG 1276
QY 1643 atttgtgtcattggtccaatcgttcccgagtggtactacatccatcaggaaggtgacc 1702
DB 1277 ATTTCGTGTCATGTCGCAATTCGTCCTCCAGTGTACTACATCTGAGAAAGTGTCC 1336
QY 1703 ttgaatgaatgactatcggtlaacccgacaaagagagaccglatcccttaacctag 1762
DB 1337 TTGATGAATATGACTATCGTATTAACCCGACCAAAAGAGAAACCCGATACCTTACCTTAG 1396
QY 1763 ctgttaatgctggtatgagagcggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1822
DB 1397 CTGTATGCTCGGATTAAGGAGCGCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1456
QY 1823 caggaacacagcagctagagaaagactgtgtgactacatccatccatcaggaagatc 1882
DB 1457 CAGGACACAGCAGCTAGAGAAAGACTTGTGTGACTACATCCGCGCATGACAGAAATC 1516
QY 1883 tccgagccttaaaagagctgtttagcaacctagaaagttccctgactcttctgtgaag 1942
DB 1517 TCCGAGCCTTAAAGAGTCTGTAGCAACCTGAAAGAGTCCCTGACTTCTTGTGTGAAG 1576
QY 1943 tggttctacagaccggagagagatgagctgtgttcttaagaaaggttgattgtg 2002
DB 1577 TGGTTCTACAGAACCGAGGGGATTAAGATCTGTCTTTCTTAAGAAAGGTGGTTATGTG 1636
QY 2003 cagccttaaaagaaatgttgccttctatgttagatcaactcagagagcactca 2062
DB 1637 CAGCCTTAATAAGAAAGTGTCTCTATGTATGATCATCTCAGAGACCTTCAAGAGCTCA 1696
QY 2063 tgaacaagcttgaagaaagttagagagcgtcgagaggaagagagagctgacaaaggtc 2122
DB 1697 TGAACAAGCTTGAAGAAAGGTTAGAGAGCGCTGCAAGAGGAAAGAGAGGCTGACCAAGGGGT 1756
QY 2123 ggttgaagaaatgttcaacagagcttcttgtagatgacacccctgtctctcagag 2182
DB 1757 GGTGTAAGAGATGTTCAACAGCTCTCTGTGATGACCAACCTGTCTCTCTGAGAGG 1816
QY 2183 ggcctcagtagtccctgtctctgttacttaagttgggcttgccttaataagttgtg 2242
DB 1817 GGCCTCTGTAGTCTGCTGCTGTACTTACAGTGTGGGCTTGTCTTATTAATAGGTTTG 1876
QY 2243 ttgccttgttagagaaagagtgagtgagtcagtcagatcagatgtttaggcaacagtagc 2302

Db 1877 TTGCTTTGTTAGACAGAGTGTGACGTCACATCTGTTACTTAGGCAACAGTACC 1936
QY 2303 aagggctctgagcccaaggaagaactgcctctag 2337
Db 1937 AAGGCTTCTGAGCCAGCAAGAACTGACTCTAG 1971

RESULT 11
AF038601 7333 bp mRNA 21-MAY-1998
LOCUS Sus scrofa porcine endogenous retrovirus ERV-PK15 mRNA, complete
DEFINITION
ACCESSION AF038601
VERSION AF038601.1 GI:3133304
KEYWORDS
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
AUTHORS Akiyoshi,D.E., Denaro,M., Zhu,H., Greenstein,J.L., Banerjee,P. and
Fishman,J.A.
TITLE Identification of a full-length cDNA for an endogenous retrovirus
of miniature swine
JOURNAL J. Virol. 72 (5), 4503-4507 (1998)
MEDLINE 98216827
REFERENCE 2 (bases 1 to 7333)
AUTHORS Fishman,J.A.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1997) Infectious Disease Unit, Massachusetts
General Hospital, 55 Fruit Street, Boston, MA 02114, USA
3 (bases 1 to 7333)
REFERENCE Fishman,J.A.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-1998) Infectious Disease Unit, Massachusetts
General Hospital, 55 Fruit Street, Boston, MA 02114, USA
REMARK
COMMENT Sequence update by submitter
On May 16, 1998 this sequence version replaced gi:3116445.
FEATURES
source
1..7333
/organism="Sus scrofa"
/strain="miniature swine"
/db_xref="taxon:9823"
/cell_line="PK-15"
/note="porcine endogenous retrovirus ERV-PK15; infective
virus but represents a naturally occurring transcript"
1..578
LTR
CDS
598..2172
/codon_start=1
/product="gag protein"
/protein_id="AAC16766.1"
/db_xref="GI:3116446"
/translation="MGQIVTTPLSLTLDHMTVRSRAHNLSVYKKGKHWQTFCASEMP
TFDVGWSEGTENSEILAVKAILFQTPGSHPODEPIILWDLADBPWPVYPMIN
KPRKRGRIILALGKRNKSHAKVESPRIYPIEPEPTWPPQVPPPAQCAVAG
PSAPGAVVVGPAAGTRSRGATPERDEIAILPLRTYGPMPGGOLOPVOYVPS
ADLVNWKTHNHPFSEDPORLTGLVLSLMSHQPTWDDCOOLTLTTEERERILIEA
KKVNGAGRPTOLONTDMGFPILTRBMDVNTAGRSKATIKQALVACIRGASRP
TNLAKVREYMGPNPPSVLEKLEAKRRTTPDPITSEAKASVALAFTGOSALDIR
KILQLEGLQAEALRDILVREAEKYIRRETEEEKERERERERERERERERERER
KTLAAVGEKSSRRERERFRKIRSGPROSGNLGNRTPLDKQCAVCEKEGHWANCPK
KGNKGPVLALEDKD"
2320..3552
/note="truncated protein"
/codon_start=1
/product="pol protein"
/protein_id="AAC16767.1"
/db_xref="GI:3133305"
/translation="MGATGROYPMTRRTVDLGVRVTHSEFVTPCPVILGRDL
TKGAQISFEGRGREVSNNKPIVTLQDLDDERLSPVQKPRQDQSNIEQPPQW
AETAGMGLAKVPPVOYIOKASATPVSRYRPLSREREGIMPVORLIDGILVPO
SPWNPPLLVRRPGTNDIRPVQDLREVKRQDIIHPIVPEYNLSLALPEERNYTVL
DLKDAFCLRLHPTISPLTFPEMRDPGTRGQDLTMTPLDQGFKNSTPIEDALHRLD

ANRIHQPVLTLLQYDDLLLAGATKODCLEGRKALLLESLDGRASAKKAOICRE
VLYLGVSLRGORMLTEARKRTVVOIIPAPPTARQVREFLESTAFCLMIPGFTLAP
LVLPLTRKGGCCLPQGGSK"
4803..6725
CDS
/note="env protein; N terminus deleted"
/codon_start=1
/pseudo
6726..7333
LTR
BASE COUNT 1984 a 1788 c 1885 g 1676 t
ORIGIN

Query Match 27.5%; Score 813; DB 4; Length 7333;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 863; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2082 gttagaagcgctgcgaaggaagagagagctgaccaggggtgtgttgaagatgttcaa 2141
Db 6470 GTTAGAGAGGCGCTGMAAGGAAGAGAGAGGCTGACAGGGGTGTTGAAGATGTTCAA 6529
QY 2142 caggtctccttgatgagaccaccctgctctctgctctgacaggggccccctagtcctgct 2201
Db 6530 CAGGCTCTCTTGATGACACCCCTGCTTCTGCTCTGACGGGGCCCTAGTAGTCTGCT 6589
QY 2202 cctgttacttaacgltgggacctgtcttaataataagtttltgcttltgttgaagacg 2261
Db 6590 CCGTACTTACAGTGTGGGCTTGCTGTTAATTAATAGTTGTGCCPTTGTATAGAACG 6649
QY 2262 agtgaagtcagttccaagatcgttacttgaagcaaacagtaacagagcctcttgagcaag 2321
Db 6650 AGTAGTGACAGTCCAGATCATGACTTATGAGCAACAGTACCAAGGCTCTGAGCCCAAG 6709
QY 2322 agaaactgaccttgaagcttccctccagcttctaagattgaactataaagaagaagagtg 2381
Db 6710 AGAAGTACCTCTTACCTCTTCCAGTTCTTAGATTAGACTATTATACAGACAGAGATG 6769
QY 2382 ggaatgaagaagatgaataatgcaactaaacctccaggaaccaggaagtttaataaag 2441
Db 6770 GGAATGAAGAGATGAATAATGCAACCTAACCTCCAGAACCCAGGAAGTTAATAAAG 6829
QY 2442 ctctaaatgccccgaattacagaccctgctgctgcagtaaatagtagtaagttaca 2501
Db 6830 CTCTAAATGCCCCCGAATTCAGACCCGCTGCTGCAGTAATAGTGAAGAGTACA 6889
QY 2502 ctctctatgttccaagggcctgctatctctgctcctaaatgaagatacaaggaatgaagttga 2561
Db 6890 CTTCCTATTGTTCCAGGGCCCTGCTATCTGCTGCTTAAGTAATACAGGAATGAGTTGA 6949
QY 2562 ctaatgccttacttgatctgtlaaaactgactgagaccatagagaattgatatcat 2621
Db 6950 CTATCGCTTATCTGAGATTCTGTAACACTGACTGGACCATAGAGAATTGATTACAT 7009
QY 2622 tgacagccctagtagacctatctcaactgcaactgttcaactctgcccagagccacgag 2681
Db 7010 TGACAGCCCTAGTAGCACTATCTCAACTGCAATGTGTCACCTGTGCCAGGACCCACG 7069
QY 2682 atggcgacctcggagagctattttaaatgatgtgtccaagagagggcgctctcgatat 2741
Db 7070 ATGGCGACCTCCGGAGCTATTTAATGATGTGTCACGAGGCGGGCTCTCGAATAT 7129
QY 2742 ttaaaatgatgtgtccatgtgagcgcgggcctcgaatatttaaatgatgtgtgtgac 2801
Db 7130 TTAATAATGATTTGTCATGAGAGCGCGGCTCTGATTTTAAATATGTTGTGTGAC 7189
QY 2802 gcacagccttgtgtgaaacccataaagctgtccgactccgagcggcgagcagt 2861
Db 7190 GCACAGCCTTGTGTGTAACCCCATTAAGACTGCGCATTCGCACTCGGAGCCGCGAGT 7249
QY 2862 ccttaccctcctgctgtgtgtagcagctgtggccccagcgagcgttggataataaatcctc 2921
Db 7250 CCTTACCCCTGCTGTGTGACAGCTGTGGCCCCAGCGCTTGGAATTAATAATCCTCT 7309
QY 2922 tgccttgcatacaaaaaaaaaa 2945

KKIQRLLEGIQENLELDVIREAKYVYRRETEPERKQREKEREERERRDRRQENKLT
KILAAVEBGRSREERERDRFRIRSGROSGLNGNTPTDKQCAACKERHMRAPCK
KNGKGVYLALEEDDGRGSDPLPEPVTLLKDEQAPVLTKEGAPVCAKESVLLQDGLK
KEKSMWAGATGROYPMWTRTVDLGVGRVYHSLVIECVYPLIGDGLILKMKAOI
SFEQGRREVSNNKPLTYLTILOLDEYVLSPOYKPODDIOSMLPEOPFOAARFAGMC
LAKOYPROVIOUKASATPVSNOYPLSEARNGIYPHVORLIQOGLIVAVOSPMTPL
LPRKPGTNDYRPVODLREVNKRVDIHPTVPNPNLISALPEPENWTVLIDLKAEF
CLRLHPTSOPLERFMDPGRGRTGLWRLOPQFNKSPRTIFEDALRDLANEFTIOH
POVLLIQYVDDLLAGATKODCESTKALLLESLQIRASAKKQICORREYTYGS
LKGOGWLEAKRTVVOIPAPPTAKOYREFGTQFRLMAGPAPATIAALPLYTEKE
KEFSNAPRHOXAPAIKRAKILISAPALIPYTKFTLYBERKVAAGVITOTLGPY
RRPVAYLSKKDPPVAGPVCILKAIANAVALIKKDKLILGONTIVYPAHLENTVRO
PPDRWMTNARMTHYOSLLTERVETAPPAVALPALLEDEPETHOCHOLILEETG
YAKDLITPLTEGVLWFTDSSYVEEKRMAGAAYVGGRTIMASRIPESISQKAE
LMALTOALRLAEGKSINITYDSRYAFATAHGAIVYKORGLTSAGRSIKKEESTLS
LEALHPKRLAIYIHCGRHAKADLISRONADRYAKAOAVNLPIIETPKPEPR
ROYTLEDMOEIKKIDOFSETPGTCYVTEKILIPKSGLEYVOQHILTHGKRHIO
QIVRTSPYHVLRLPGVADSVYKHCYPCOLVANNBSRIPGKRRLRSGHAMEVDTE
VAPRAYGNKYLLVFDTSQWVEAPYTKKETSTVAKKILEETLPPRFQIPVIGSDNG
PAFVAOSQGLAKILGIDMKLHCAYRPOSSQOVERMNTIKETILKLTTEGINDMA
LLPEVLFRVNTPGQGLPYELLXGPPYVEIASVADVLISQPLSRILKALEVY
ROKAMKOLREAYSGEDLOVPHRFQVGSVYVRRHACNLETRMKGPVLYLLTPTAY
KYERIPWTHASHVKBAPPDPSGMAKETEENPKIRLHRVYPSVNNSSMHPITSR
HLPIRGKPKRLKIPLSFASIAWFLTSLITPOVNGKRLVDSNHSKPLSTWLLTDSG
TGININISQGEAPLGTWPELIVCLRSVIFGLNDATPPDVLRAFGVYCGPPNNE
YCGNPODEFCKOMSCITSNDGNMKWPVSOODRVSYPFANNPTSYOMFYGGKMKMG
OVOKDVRNKQKISLDLDYIKTSETKGEONKOKNGVITSMGIVYGGGKMKMG
VLTIRLETOMEPPVAYIGPNKGLAEOGPLOEOBPSPNDYNTTSGSVTEPNITL
KTGAKPLSLIOGAPALNSTTPEATSSWCLASPPYEGMARCKAVTKKEHDCC
TWGSONKLLTEVSGKSTICIGVPPSHOHLNHEAFNRTSESOTLVYGVYRWMACT
GLTPCVSTLVENOTKDFVMVOIYPRVYVPEKALDYRVRNPKREPISLTLAV
LGLGVACVGTGTALITGPOOLLEKRLIVNRYIREDOLAEKSYNLEESITLSEV
VLONRGDLLEFLKEGCLVALKECCCYVHSGAIRDMSKLRERLERRREREADQ
GMPEGMFNRSPMPTLLSALNGPLVLLILLTVGCILNRPVAPRERVSAYQIMVLR
QOYGLISQGERDL"

mat_peptide
2690..6271
/gene="pol"
/function="retroviral protease and polymerase"
/product="pol protein"

gene
2690..6271
/gene="pol"
/gene="env"
/function="retroviral envelope"
/product="Env protein"

gene
6150..8129
/gene="env"
/gene="env"
/gene="env"

LTR
8182..8849
8742..8747

polyA_signal
2355 a 2191 c 2281 g 2022 t

BASE COUNT
2355 a 2191 c 2281 g 2022 t

ORIGIN

Query Match 21.2%; Score 626; DB 14; Length 8849;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 676; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2082 gttagaagagcgtcgaaggaagagagcgtgacacaggggtgtttgaagatgttcaa 2141
|||||
Db 7877 gttagaagagcgtcgaaggaaggaagagcgtgacacaggggtgtttgaagatgttcaa 7936
QY 2142 caggtcctctgatacaccacccctgtctctgctgacaggggcccctgtagtctgtc 2201
|||||
Db 7937 caggtcctctgatacaccacccctgtctctgctgacaggggcccctgtagtctgtc 7996
QY 2202 cctgttactacagttggcctgtcttaataagttgttgccttgttagaagacg 2261
|||||
Db 7997 cctgttactacagttggcctgtcttaataagttgttgccttgttagaagacg 8056
QY 2262 agtgaatgacgtccagatactgtagcaacagttacaaagccttctgaagcaag 2321
|||||
Db 8057 agtgaatgacgtccagatactgtagcaacagttacaaagccttctgaagcaag 8116

QY 2322 agaactctacctctgagccttccagcttctaagattagaactatlaacaagaagaagtg 2391
|||||
Db 8117 AGAACTACACCTCTAGCCTTCCAGTCTAGATAGAACTATTAAACAAGACAAGAACTG 8176
QY 2382 gggatgaaagatgaaataatgcaacttaacccctccagaacccaggaagttaataaag 2441
|||||
Db 8177 GGAATGAAGAAGATGAAAATGCAACCTAACCTCCCAACCCAGGAAGTTAAATAAAG 8236
QY 2442 ccttaatggccccgaatattacagacccctgctgctgagcaagtaaatagtagaagtgaca 2501
|||||
Db 8237 CTTAAATGCCCCCAATTTCCAGACCTGCTGCTGCCAGTAATAGTTAAGGTCCACA 8296
QY 2502 ctctcatgtgtccagggcctgtcatacctggccttaagtaagaatacaggaatgagtga 2561
|||||
Db 8297 CTTCTATATTGTTCCAGGGCCTGCTATCTCTGGGCTTAAGTAAGATAACAGAAATGAGTTGA 8356
QY 2562 ctaatgccttatctgtagtctgttaaacatgtagcagccatagaagaattgattacacat 2621
|||||
Db 8357 CTAAATCGCTTATCTGATTTCTGTAAACTGACTGACCATAGAAAGAAATTGATTACACAT 8416
QY 2622 tgacagccctagtgacactatactcaactgcaatctgtcaactctgccccagagccagag 2681
|||||
Db 8417 TGACAGCCCTTAGTACTATCTCACTGCACTGATCTGTCACTGTGCCAGAGCCACGACG 8476
QY 2682 atgcgacctccgagatcttlaaataatgtagtgcacagagcgagcgtctcgatatt 2741
|||||
Db 8477 ATCGGACACTCCGGAGCATTTTAAATGATTGTGCCAGGAGCGGCGCTGTGATATT 8536
QY 2742 ttaaaatgattgtgtcca 2758
|||||
Db 8537 TTAATGATTGTGTCCA 8553

RESULT 14

AX002802
LOCUS AX002802 2462 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 1 from Patent WO9853104.
ACCESSION AX002802
VERSION AX002802.1 GI:9885131
KEYWORDS

SOURCE
ORGANISM
porcine endogenous retrovirus.
porcine endogenous retrovirus
Viruses; Retroid viruses; Retroviridae; Mammalian type C
retroviruses; 1-Mammalian type C virus group.

REFERENCE

1 (bases 1 to 2462)
Stoye,J.P. and Weiss,R.A.
Detection of retroviral subtypes based upon envelope specific
sequences
Patent: WO 9853104-A 1 26-NOV-1998;

JOURNAL

MEDICAL RES COUNCIL (GB); STOYE JONATHAN PAUL (GB); WEISS ROBIN
ANTHONY (GB)

FEATURES

source

1..2462
/organism="porcine endogenous retrovirus"
/db_xref="taxon:61673"

BASE COUNT

715 a 573 c 575 g 599 t

ORIGIN

Query Match 15.5%; Score 457; DB 6; Length 2462;
Best Local Similarity 99.8%; Pred. No. 3.6e-252;
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2082 gttagaagagcgtcgaaggaagagagcgtgacacaggggtgtttgaagatgttcaa 2141
|||||
Db 1938 gttagaagagcgtcgaaggaaggaagagcgtgacacaggggtgtttgaagatgttcaa 1997
QY 2142 caggtcctctgatacaccacccctgtctctgctgacaggggcccctgtagtctgtc 2201
|||||
Db 1998 caggtcctctgatacaccacccctgtctctgctgacaggggcccctgtagtctgtc 2057
QY 2202 cctgttactacagttggcctgtcttaataagttgttgccttgttagaagacg 2261
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on:      February 24, 2002, 07:34:23 ; Search time 584.4 Seconds
              (without alignments)
              12042.750 Million cell updates/sec
```

Title: US-09-171-553B-3

Perfect score: 8209
Sequence: 1 gtggtgtacgactgtggcc.....aaaaaaaaaaaaaaaaaa 8209

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched:

930621 seqs, 428662619 residues

Word size :

Total number of hits satisfying chosen parameters

1861242

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database :

1: /SID22/gcgdata/geneseq/geneseqn/NA1580.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn/NA1581.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn/NA1582.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn/NA1583.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn/NA1585.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn/NA1585.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn/NA1586.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn/NA1587.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn/NA1588.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn/NA1589.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn/NA1590.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn/NA1591.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn/NA1592.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn/NA1593.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn/NA1594.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn/NA1595.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn/NA1596.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn/NA1597.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn/NA1598.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn/NA1599.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8209	100.0	8209	18	AAV09700	Porcine retrovirus
2	7802	95.0	8196	18	AAV09659	Porcine retrovirus
3	2967	36.1	7333	22	AAE77726	Defective retrovirus
4	2967	36.1	7393	18	AAE77483	Porcine retrovirus
5	2682	32.3	3320	18	AAV09698	Porcine retrovirus
6	1499	18.3	1974	18	AAV09703	Porcine retrovirus
7	1466	17.9	3482	22	AAV82149	ply endogenous retri
8	848	10.3	4402	22	AAV67020	PERV env protein C
9	797	9.7	6076	22	AAV67021	PERV env protein C
10	797	9.7	7362	22	AAV67019	PERV env protein C
11	746	9.1	6076	22	AAV67032	PERV env protein C

C	4	5	29	0.4	8202	17	AAT90280
	44	29	0.4	7308	18	AAT90688	
	43	29	0.4	7308	18	AAT90688	
	42	29	0.4	7308	18	AAT90688	
	41	29	0.4	5178	21	AAT90688	
	40	29	0.4	3643	21	AAT90688	
	39	29	0.4	3557	7	AAT90680	
	38	29	0.4	1234	18	AAT904005	
	37	30	0.4	1431	21	AAT907334	
	36	31	0.4	2702	19	AAT955872	
	35	34	0.4	1980	22	AAC67018	
	34	34	0.4	600	22	AAC67018	
	33	34	0.4	600	22	AAT962575	
	32	34	0.4	600	22	AAT962522	
	31	34	0.4	600	22	AAT962522	
	30	34	0.4	599	22	AAT962525	
	29	34	0.4	599	22	AAT962525	
	28	34	0.4	599	22	AAT962525	
	27	34	0.4	599	22	AAT962525	
	26	34	0.4	599	22	AAT962525	
	25	34	0.4	599	22	AAT962525	
	24	35	0.4	8655	20	AAT969750	
	23	35	0.4	127	22	AAC6800225	
	22	41	0.5	45	22	AAC82275	
	21	45	0.5	45	22	AAC82275	
	20	91	1.1	704	22	AAC67006	
	19	91	1.1	633	22	AAC67007	
	18	375	4.6	8060	18	AAT9778255	
	17	375	4.6	8060	18	AAT9778255	
	16	457	5.6	2462	20	AA9B827043	
	15	496	6.0	7873	22	AA9B827043	
	14	570	6.9	4918	22	AAC67022	
	13	704	8.6	8132	22	AAT977727	
	12	704	8.6	7892	18	AAT94884	

ALIGNMENTS

Nucleotide sequence
PERV env protein c
PERV env protein c
Pig endogenous ret
Porcine retrovirus
Tsukuba-1 CDNA. P
pREV-MSL LTR SEQ I
I Tamal LTR SEQ ID N
Human retrovirus D
Human retrovirus D
DDEEV RT sequence,
Mus damni endogeno
Pig endogenous ret
Pig endogenous ret
Pig endogenous ret
Pig endogenous ret
Pig endogenous ret
Pig endogenous ret
Pig endogenous ret
PERV-1-15 env prot
Plant acetylacetate
Human ORX ORF2889.
Retrovirus 4070A p
Human pro-growth h
MTV reverse transcr
Gibbon leukemia vi
Plasmid HCMV-intro
Plasmid HCMV-intro
plasmid CB6 coding
Novel AMP/MCF vtru

RESULT	1
AAV09700	
ID	AAV09700 standard; DNA; 8209 BP.
AC	
XX	AAV09700;
DT	20-MAY-1998 (first entry)
XX	
XX	Porcine retrovirus DNA encoding the LTR and GAG, POL and ENV proteins.
DE	
XX	Porcine retrovirus; POEV; POL protein; ENV protein; GAG protein;
KM	virion core polypeptide; polymerase protein; envelope protein;
KW	vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds
XX	
OS	Porcine retrovirus.
XX	
Key	Location/Qualifiers
FT	1..61
FT	/*tag- a
FT	/note="R LTR domain"
FT	1..588
FT	/*tag- b
FT	62..143
FT	/*tag- c
FT	/note="U5 LTR domain"
FT	588..2162
FT	/*tag- d
FT	/product= GAG protein
FT	/note="Viral core polypeptide"
FT	2163..5747
FT	/*tag- e
FT	/product= POL protein
FT	/note="polymerase peptide"
FT	5620..7590
CDS	

PR	10-FEB-1997 ;	97GB-0002668.
PR	19-APR-1996 ;	96GB-0008164.
XX		
XX		
PA	(IMOT-) IMOTRAN LTD.	
PA	(QONE-) Q-ONE BIOTECH LTD.	
XX		
XX		
PI	Galbraith DN, Haworth C, Lees GM, Smith KT;	
DR	WPI: 1997-535851/49.	
DR	P-PSDB; AAM39271; AAM39272; AAM39273.	
XX		
XX		
PT	Poly nucleotide encoding porcine retrovirus expression product -	
PT	useful to develop products for use in vaccines, diagnosis and	
PT	xeno-transplantation	
XX		
XX		
PS	Claim 4; Fig 3; 69pp; English.	
CC		
CC	This DNA sequence encodes the porcine retrovirus (POEV) virion core	
CC	polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins and	
CC	also includes the Long Terminal Repeat (LTR). These proteins can be used	
CC	to develop viral vaccines, antisense nucleic acids, ribozymes and other	
CC	antiviral agents. They can also be used in xeno-transplantation	
CC	technology and as diagnostic tools.	
XX		
SO	Sequence 8209 BP; 2168 A; 2064 C; 2154 G; 1823 T; 0 other;	

Query Match	Similarity	100.0%	Score 8209;	DB 18;	Length 8209;
Best Local	Similarity	100.0%	Pred. No. 0;		
Matches 8209;	Conservative	0;	Mismatches	0;	Gaps 0;
QY	1	gtggtgtaacgaactgtagggccccacgacgcgtcttgaaataaaatccctctgctgtttgcac	60		
Db	1	gtggtgtaacgaactgtagggccccacgacgcgtcttgaaataaaatccctctgctgtttgcac	60		
QY	61	aagacgcgtctctgtgagtgatcttgagggtgttcgcctcttcogagccggagagggagat	120		
Db	61	aagacgcgtctctctgagtgatcttgagggtgttcgcctcttcogagccggagagggagat	120		
QY	121	tgtttcttaactgagccttcattcttggtgtcgtttgscggagaaatctcttgagccaccctta	180		
Db	121	tgtttcttaactgagccttcattcttggtgtcgtttgscggagaaatctcttgagccaccctta	180		
QY	181	caaccgagaaacgcgaactgtagagtgaaagagatccctcttgagaaagtgtagtgctgcgac	240		
Db	181	caaccgagaaacgcgaactgtagagtgaaagagatccctcttgagaaagtgtagtgctgcgac	240		
QY	241	ggcgctctctgtcttgagtgctgtttcttggtgtagtgacgacgtcttcggtttgcagctgcct	300		
Db	241	ggcgctctctgtcttgagtgctgtttcttggtgtagtgacgacgtcttcggtttgcagctgcct	300		
QY	301	ctcagacccgtaagagaaacttgagagactgtagatcatagcagaaagtgctaaagagatcaacagagctg	360		
Db	301	ctcagacccgtaagagaaacttgagagactgtagatcatagcagaaagtgctaaagagatcaacagagctg	360		
QY	361	ccaacccttgaggagacgcccccgagaggttgaggagagccagagagacgcctgtgtctctact	420		
Db	361	ccaacccttgaggagacgcccccgagaggttgaggagagccagagagacgcctgtgtctctact	420		
QY	421	gtcggtcaagagagacgagttctgtgtttgaaagcgaaagctcccccctcgcgcgcgtccg	480		
Db	421	gtcggtcaagagagacgagttctgtgtttgaaagcgaaagctcccccctcgcgcgcgtccg	480		
QY	481	actcttttgcctgtcttgtagaagacgagaggggtctcggtgtgtctgtatctgttgggtt	540		
Db	481	actcttttgcctgtcttgtagaagacgagaggggtctcggtgtgtctgtatctgttgggtt	540		
QY	541	ctgtctcgtgtgtctcttgctcttgctgcgctccctgtgtctacagttttaatatgtagaagacag	600		
Db	541	ctgtctcgtgtgtctcttgctcttgctgcgctccctgtgtctacagttttaatatgtagaagacag	600		
QY	601	tgaatcccccccttagtttgactctgcacatctggaactggaagttagatccaaggctcata	660		

Db 601 tgactcccccttagttgactctgcacatctgactgaatgaatgataccaggtccata 660
Qy 661 attgtcagttcaggttaagaaaggaacttgcagacttctgtccctgaatggccaa 720
Qy 661 attgtcagttcaggttaagaaaggaacttgcagacttctgtccctgaatggccaa 720
Qy 721 cattcgatgttgatgacatcagaggggacattatctgaaatatactctgcttga 780
Db 721 cattcgatgttgatgacatcagaggggacattatctgaaatatactctgcttga 780
Qy 781 aggaacatcatlcttcagacttgaccccggtctcatcccgatcaagacccatactta 840
Db 781 aggaacatcatlcttcagacttgaccccggtctcatcccgatcaagacccatactta 840
Qy 841 cgtggaagatcttgacgaagatctccgcacatgggttaaacatctgcttaataacca 900
Db 841 cgtggaagatcttgacgaagatctccgcacatgggttaaacatctgcttaataacca 900
Qy 901 gaaagccaggtcccccgaatccctgtctcttgagagagaacaaacacatccgcgaagaa 960
Db 901 gaaagccaggtcccccgaatccctgtctcttgagagagaacaaacacatccgcgaagaa 960
Qy 961 tccagccctcttctcgtatctaccccgagatcgagagagccgcgacttgcgcgaaccc 1020
Db 961 tccagccctcttctcgtatctaccccgagatcgagagagccgcgacttgcgcgaaccc 1020
Qy 1021 caactgttcccccacccccccttaccagcacaggtgtctgtagagagacactctgcctc 1080
Db 1021 caactgttcccccacccccccttaccagcacaggtgtctgtagagagacactctgcctc 1080
Qy 1081 ctgagagctccgtgtgtagagagacactctgtagagagacactctgtagagagacact 1140
Db 1081 ctgagagctccgtgtgtagagagacactctgtagagagacactctgtagagagacact 1140
Qy 1141 cggagcgagacagagagatcgagatataccgctgtagagacactctgtagagagacact 1200
Db 1141 cggagcgagacagagagatcgagatataccgctgtagagacactctgtagagagacact 1200
Qy 1201 ggggccaattgcagccccctccagatctgagacttcttctgcagatctctataatgga 1260
Db 1201 ggggccaattgcagccccctccagatctgagacttcttctgcagatctctataatgga 1260
Qy 1261 aactaaacatcccccttctcggagagatcccccaagctccacggggtgtgtagagatcc 1320
Db 1261 aactaaacatcccccttctcggagagatcccccaagctccacggggtgtgtagagatcc 1320
Qy 1321 ttatgttctctacacagctacttgagatgtgtcaacagctgtctgcagacactctca 1380
Db 1321 ttatgttctctacacagctacttgagatgtgtcaacagctgtctgcagacactctca 1380
Qy 1381 caacgagagagcgagagagaaatctctgttagaggttagaataatctctgcgagcgag 1440
Db 1381 caacgagagagcgagagagaaatctctgttagaggttagaataatctctgcgagcgag 1440
Qy 1441 ggcgacccaagctgtgtaaaatctgacatggatctccctgactgcgccccgtt 1500
Db 1441 ggcgacccaagctgtgtaaaatctgacatggatctccctgactgcgccccgtt 1500
Qy 1501 gggactacaacacgctgaaaggttagggagagctgtaaaatctctgcgagcgctgtgtg 1560
Db 1501 gggactacaacacgctgaaaggttagggagagctgtaaaatctctgcgagcgctgtgtg 1560
Qy 1561 cgggtctccggggcgccctaaagacggccactaatctgctaaggttagagaggtgagtc 1620
Db 1561 cgggtctccggggcgccctaaagacggccactaatctgctaaggttagagaggtgagtc 1620
Qy 1621 agggacgaaagacactccctcggtatcttctgtagaggtctatgaaagccttcagcggt 1680
Db 1621 agggacgaaagacactccctcggtatcttctgtagaggtctatgaaagccttcagcggt 1680
Qy 1681 tcaccccttctgactacactcagagagccagaaagcctcagtgccctgcttcatctg 1740
Db 1681 tcaccccttctgactacactcagagagccagaaagcctcagtgccctgcttcatctg 1740

Qy 1741 ggcagctgcgtctggaatcagagaaacactcagagacttgaaaggttaccagaggtg 1800
Db 1741 ggcagctgcgtctggaatcagagaaacactcagagacttgaaaggttaccagaggtg 1800
Qy 1801 agttacgttgatctagtgagagagcgagagaggtctatctacagaaagggagacaaagag 1860
Db 1801 agttacgttgatctagtgagagagcgagagaggtctatctacagaaagggagacaaagag 1860
Qy 1861 agaaagaaacagagaaagaaagagagagaaagagagagagagagagagagagagagag 1920
Db 1861 agaaagaaacagagaaagaaagagagagagagagagagagagagagagagagagagag 1920
Qy 1921 aagaaagaaatcttgcacaaagatcttgcagaggtgtgtaagggagagagagagagag 1980
Db 1921 aagaaagaaatcttgcacaaagatcttgcagaggtgtgtaagggagagagagagagag 1980
Qy 1981 gagagagagatcttgaagaaatctgtagagccctgaagccctgaagagagagagagag 2040
Db 1981 gagagagagatcttgaagaaatctgtagagccctgaagccctgaagagagagagagag 2040
Qy 2041 ggaacccactcgaagagagacaggtgtgtagatctgtaagaaagagacactgagcaag 2100
Db 2041 ggaacccactcgaagagagacaggtgtgtagatctgtaagaaagagacactgagcaag 2100
Qy 2101 actgcccacag 2160
Db 2101 actgcccacag 2160
Qy 2161 agggag 2220
Db 2161 agggag 2220
Qy 2221 aaccagcttgatctcgtgtgtgtagatccggagagagagagagagagagagagagag 2280
Db 2221 aaccagcttgatctcgtgtgtgtagatccggagagagagagagagagagagagagag 2280
Qy 2281 gaaacactaaagaaagaaagaaatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2340
Db 2281 gaaacactaaagaaagaaagaaatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2340
Qy 2341 ggaactaccgaaagagacggtgactgtgagaggtgagagagagagagagagagagag 2400
Db 2341 ggaactaccgaaagagacggtgactgtgagaggtgagagagagagagagagagagag 2400
Qy 2401 tcccgaggtgcccagtaacccctctcagtagagagacttaccagagatggagagctcaaa 2460
Db 2401 tcccgaggtgcccagtaacccctctcagtagagagacttaccagagatggagagctcaaa 2460
Qy 2461 ttctcttgaacaaagagagacagagatgtctgtgtaataacaaacccatcaactgtgtga 2520
Db 2461 ttctcttgaacaaagagagacagagatgtctgtgtaataacaaacccatcaactgtgtga 2520
Qy 2521 cccctcaatataatgataatgataatcgaactatacttccccaagtaaaccttgataaagata 2580
Db 2521 cccctcaatataatgataatgataatcgaactatacttccccaagtaaaccttgataaagata 2580
Qy 2581 tacaagctgt 2640
Db 2581 tacaagctgt 2640
Qy 2641 caaagacaggttcccccacaggtatctcaactgaagccagtgctacacacatcaagca 2700
Db 2641 caaagacaggttcccccacaggtatctcaactgaagccagtgctacacacatcaagca 2700
Qy 2701 gacagtaaccccttgagtagagagagctccggagagagagagagagagagagagagag 2760
Db 2701 gacagtaaccccttgagtagagagagctccggagagagagagagagagagagagagag 2760
Qy 2761 tccaaagagagatctcgt 2820
Db 2761 tccaaagagagatctcgt 2820

QY 2821 ggaagcctggagccaatgatattatgcacgaatgacagagactgagagagatcacaataaagg 2880
DB 2821 ggaagcctggagccaatgatattatgcacgaatgacagagactgagagagatcacaataaagg 2880
QY 2881 tgcagagacataacccaacgggtcccgaaacctataacctcttgagcgccctccgctg 2940
DB 2881 tgcagagacataacccaacgggtcccgaaacctataacctcttgagcgccctccgctg 2940
QY 2941 aacggaactgttacacagatattgacttaaaagatgctctctctgctgagattacacc 3000
DB 2941 aacggaactgttacacagatattgacttaaaagatgctctctctgctgagattacacc 3000
QY 3001 ccaactagccaacacacttttgccttcgagatgagagatcccaaggtacaggaagacgggc 3060
DB 3001 ccaactagccaacacacttttgccttcgagatgagagatcccaaggtacaggaagacgggc 3060
QY 3061 agctcacctggagcccgacgcccccaagggttcaagaactcccgacacactcttgacgaag 3120
DB 3061 agctcacctggagcccgacgcccccaagggttcaagaactcccgacacactcttgacgaag 3120
QY 3121 cccctacacagagaccttggccaactcagagatcccaacccctcaggtgacccctccacgt 3180
DB 3121 cccctacacagagaccttggccaactcagagatcccaacccctcaggtgacccctccacgt 3180
QY 3181 aagtgagatgacctgtcttgcgagagacccaacacagagactgttagaaggtacagaa 3240
DB 3181 aagtgagatgacctgtcttgcgagagacccaacacagagactgttagaaggtacagaa 3240
QY 3241 cactactgtggaattgtctgacactcaggtacagagcctctgtcaagaagcccaagatt 3300
DB 3241 cactactgtggaattgtctgacactcaggtacagagcctctgtcaagaagcccaagatt 3300
QY 3301 gcaagagagaggtacacttctgggttaacagtttgcgagcgagacagatgagcgagag 3360
DB 3301 gcaagagagaggtacacttctgggttaacagtttgcgagcgagacagatgagcgagag 3360
QY 3361 aggcacggagaaacactgtatgcacagataccgcccccaacacagccaacaaatgagag 3420
DB 3361 aggcacggagaaacactgtatgcacagataccgcccccaacacagccaacaaatgagag 3420
QY 3421 agttttgggagagactgtattcttcagagactgtgagatccgggtttgagacattagcag 3480
DB 3421 agttttgggagagactgtattcttcagagactgtgagatccgggtttgagacattagcag 3480
QY 3481 ccccaactacccgttaacccaagaaagggagattctcctgagctctgagccaacaga 3540
DB 3481 ccccaactacccgttaacccaagaaagggagattctcctgagctctgagccaacaga 3540
QY 3541 aggcattgagatctacaagagccctgtcagcagcctgctctgagccctccctgagc 3600
DB 3541 aggcattgagatctacaagagccctgtcagcagcctgctctgagccctccctgagc 3600
QY 3601 taactaaaccccttaaccccttaattgtgagatgagcgttaagggagtagccgagagattaa 3660
DB 3601 taactaaaccccttaaccccttaattgtgagatgagcgttaagggagtagccgagagattaa 3660
QY 3661 cccaaccccttaagacatgagagagacctgttgcactacgttccaagaagactgtatcc 3720
DB 3661 cccaaccccttaagacatgagagagacctgttgcactacgttccaagaagactgtatcc 3720
QY 3721 taagcagatgttgcagatgtctgaaagactatcgagcgtgtgagccaatactgtaag 3780
DB 3721 taagcagatgttgcagatgtctgaaagactatcgagcgtgtgagccaatactgtaag 3780
QY 3781 agcgtgaacaatgactttggagacagaaatataactgtatagcccccatgcatgtgaga 3840
DB 3781 agcgtgaacaatgactttggagacagaaatataactgtatagcccccatgcatgtgaga 3840
QY 3841 acatcgttcgagagcccaagacagatgagacccaacgcccagatgacccaatatacaa 3900
DB 3841 acatcgttcgagagcccaagacagatgagacccaacgcccagatgacccaatatacaa 3900
QY 3901 gctcgtctctcaagagaggtcacttctgcacacagcgctctcaacactgtcacatc 3960
DB 3901 gctcgtctctcaagagaggtcacttctgcacacagcgctctcaacactgtcacatc 3960

DB 3901 gctcgtctctcaagagaggtcacttctgcacacagcgctctcaacactgtcacatc 3960
QY 3961 tctcgtctcaagagacatgacacagatgacacagatgacacatgacacatgacacatgacac 4020
DB 3961 tctcgtctcaagagacatgacacagatgacacagatgacacatgacacatgacacatgacac 4020
QY 4021 agacgtgggttcgagagacacttaacacacacacacacacacacacacacacacacacacac 4080
DB 4021 agacgtgggttcgagagacacttaacacacacacacacacacacacacacacacacacacac 4080
QY 4081 tcaactgacagagacagatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 4140
DB 4081 tcaactgacagagacagatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 4140
QY 4141 agggagcccgagacatctgggccaagccttgcagagccttgcagagagacattcagccaagcctg 4200
DB 4141 agggagcccgagacatctgggccaagccttgcagagccttgcagagagacattcagccaagcctg 4200
QY 4201 agctcatgacctcagcagagacttgcagccttgcagagccttgcagagagacattcagccaagcctg 4260
DB 4201 agctcatgacctcagcagagacttgcagccttgcagagccttgcagagagacattcagccaagcctg 4260
QY 4261 cggagagcgagatgagcttgcagccttgcagccttgcagagccttgcagagagacattcagccaagcctg 4320
DB 4261 cggagagcgagatgagcttgcagccttgcagccttgcagagccttgcagagagacattcagccaagcctg 4320
QY 4321 ggttgctacacacacag 4380
DB 4321 ggttgctacacacacag 4380
QY 4381 aagccttcaatcttgcac 4440
DB 4381 aagccttcaatcttgcac 4440
QY 4441 aagatctacatctcag 4500
DB 4441 aagatctacatctcag 4500
QY 4501 ctgttcaacactcctgcac 4560
DB 4501 ctgttcaacactcctgcac 4560
QY 4561 cccctagagagcttgcag 4620
DB 4561 cccctagagagcttgcag 4620
QY 4621 cctgtatcactcctacatgag 4680
DB 4621 cctgtatcactcctacatgag 4680
QY 4681 aacagatatactcctacatgag 4740
DB 4681 aacagatatactcctacatgag 4740
QY 4741 ccccttcaatgagtgag 4800
DB 4741 ccccttcaatgagtgag 4800
QY 4801 cctgcagagctggttaatgagtgag 4860
DB 4801 cctgcagagctggttaatgagtgag 4860
QY 4861 gccaacacagcgtcagcag 4920
DB 4861 gccaacacagcgtcagcag 4920
QY 4921 acaaatatcatgagtgag 4980
DB 4921 acaaatatcatgagtgag 4980
QY 4981 agaaagagacttcaacagtgagtgag 5040
DB 4981 agaaagagacttcaacagtgagtgag 5040

QY	193	gactctgagagtaagaaggatccctcttggaaacgtgtgtgtgtctcgccggtctctgtt	252
DB	181	gactctgagagtaagaaggatccctcttggaaacgtgtgtgtgtctcgccggtctctgtt	240
QY	253	ctgagatctgtcttctcggtgtatgtgagccttctcggtttgagagctgtctctcagacgttaa	312
DB	241	ctgagatctgtcttctcggtgtatgtgagccttctcggtttgagagctgtctctcagacgttaa	300
QY	313	ggaactgagagacctgtatcaacgaaacgtgtcactagaagatacaacagctgtccacccttgggg	372
DB	301	ggaactgagagacctgtatcaacgaaacgtgtcactagaagatacaacagctgtccacccttgggg	360
QY	373	acgcccacggaggtgtggtgagagccagggagacgcctggtgtctctcaatctgtccgttcaagag	432
DB	361	acgcccacggaggtgtggtgagagccagggagacgcctggtgtctctcaatctgtccgttcaagag	420
QY	433	accgagttctgtgtcttgaagcgaanaagcttccctcccgccggtccgactcttttggct	492
DB	421	accgagttctgtgtcttgaagcgaanaagcttccctcccgccggtccgactcttttggct	480
QY	493	gctctgtgaaagacgcggagacgggttcgcctgtgtctcagatctgtgttctgtctcgtgtg	552
DB	481	gctctgtgaaagacgcggagacgggttcgcctgtgtctcagatctgtgttctgtctcgtgtg	540
QY	553	tctttgtctgtgtgtctctgtctctacgtttaaataatgagacagacagtgtacatccccc	612
DB	541	tctttgtctgtgtgtctctgtctctacgtttaaataatgagacagacagtgtacatccccc	600
QY	613	ttaagttgactctcgaccatttgttgaactgtgaagttagaatccagggctcaatatgttgaatttc	672
DB	601	ttaagttgactctcgaccatttgttgaactgtgaagttagaatccagggctcaatatgttgaatttc	660
QY	673	aggttaagaaaggagacctgtgcagactcttctgtgctctgtaaatgtgcacaatctcgatgttg	732
DB	661	aggttaagaaaggagacctgtgcagactcttctgtgctctgtaaatgtgcacaatctcgatgttg	720
QY	733	gatgtgcatcaagaggggaacctttaatcttctgaaattatccctgtgctgttgaagcaatcat	792
DB	721	gatgtgcatcaagaggggaacctttaatcttctgaaattatccctgtgctgttgaagcaatcat	780
QY	793	ttcagagctggaacccgcgctcttcaatccctgatcaagaagccctatataccttaacgtgtgcaagatt	852
DB	781	ttcagagctggaacccgcgctcttcaatccctgatcaagaagccctatataccttaacgtgtgcaagatt	840
QY	853	tgtcagaagatctctccgcgcgatgtgttaaaccaatggtcctaaataaacaaagaagccaggttc	912
DB	841	tgtcagaagatctctccgcgcgatgtgttaaaccaatggtcctaaataaacaaagaagccaggttc	900
QY	913	cccggaatccctgtgctcttctgtgagagaaacaaacacatctgcgcggaanaagctcgagccctctt	972
DB	901	cccggaatccctgtgctcttctgtgagagaaacaaacacatctgcgcggaanaagctcgagccctctt	960
QY	973	cctctgtatctaaccccgagatctcgagagagccgcgcgactgtgcccggaaaccccaacctgttccc	1032
DB	961	cctctgtatctaaccccgagatctcgagagagccgcgcgactgtgcccggaaaccccaacctgttccc	1020
QY	1033	ccacccccccttataccagacaaaggtgtctgttggagggagacctctgcgccctctctgtgagcttcgg	1092
DB	1021	ccacccccccttataccagacaaaggtgtctgttggagggagacctctgcgccctctctgtgagcttcgg	1080
QY	1093	tgtgtgagagagacctgtgcgcggagactcgagagccggagagagcgcgccaccccgagagggagag	1152
DB	1081	tgtgtgagagagacctgtgcgcggagactcgagagccggagagagcgcgccaccccgagagggagag	1140
QY	1153	acgaagatcgagatatataccgcctgtgcgcacactatgtgcctccccaatgccaagggggccaattgtc	1212
DB	1141	acgaagatcgagatatataccgcctgtgcgcacactatgtgcctccccaatgccaagggggccaattgtc	1200
QY	1213	agcccccctcagatatgtgccccttctctctgtgagatcctctataattgtgaaactaacctc	1272
DB	1201	agcccccctcagatatgtgccccttctctctgtgagatcctctataattgtgaaactaacctc	1260
QY	1273	cccccctctctcgagagatccccaacgcctcaacgggtgtgtgtgagctcccttatgttctctc	1332

[illegible]

Db 2339 gaaccgttgacttggagatgtagaagcgttaaccacatcgcttctgtgtcatcccttgatgcc 2398
QY 2413 cagtaaccctcttagtagaagacttaactgacaaagatgtagagccaaattcttttgaac 2472
Db 2399 cagtaaccctctctagtagaagacttaactgacaaagatgtagagccaaattcttttgaac 2458
QY 2473 aaggaagaccgaaggtgctgtgtaataaacaacacactaactgtgttgaacccctcaatag 2532
Db 2459 aaggaagaccgaaggtgctgtgtaataaacaacacactaactgtgttgaacccctcaatag 2518
QY 2533 atgataataatcgaactaatactccccaagtaagacctgtagaagtatacagctccgtc 2592
Db 2519 atgataataatcgaactaatactccccaagtaagacctgtagaagtatacagctccgtc 2578
QY 2593 tggagcaggtttccccaagccttggcagaaacgcagaggaatgggttggcagaagcaagttc 2652
Db 2579 tggagcaggtttccccaagccttggcagaaacgcagaggaatgggttggcagaagcaagttc 2638
QY 2653 ccccaacaggttataactgaagcaggtgtcaacacagatatacagtaacagtaacccct 2712
Db 2639 ccccaacaggttataactgaagcaggtgtcaacacagatatacagtaacagtaacccct 2698
QY 2713 tgggttagagaggtcgcagagaagaaattggcgcagatctcaaaagataatcccaacagggca 2772
Db 2699 tgggttagagaggtcgcagagaagaaattggcgcagatctcaaaagataatcccaacagggca 2758
QY 2773 tcttaattccctgtcccaatcccttggaaatcccttgcctcaggttgaagaagccttgga 2832
Db 2759 tcttaattccctgtcccaatcccttggaaatcccttgcctcaggttgaagaagccttgga 2818
QY 2833 ccaatgattatatacgaacagtaacagacttgagagaggtcaataaagaaggtgcagagacac 2892
Db 2819 ccaatgattatatacgaacagtaacagacttgagagaggtcaataaagaaggtgcagagacac 2878
QY 2893 acccaacaggtccccaacacctataacaccttgagagcgcctccgcgtgaacggaagctgt 2952
Db 2879 acccaacaggtccccaacacctataacaccttgagagcgcctccgcgtgaacggaagctgt 2938
QY 2953 acaaggtttagacttaaaagaatgcttcttgcctgagatatacaccacacacacacacac 3012
Db 2939 acaaggtttagacttaaaagaatgcttcttgcctgagatatacaccacacacacacacac 2998
QY 3013 caattttagcttcgaatgtagagatatacaggtatccaggtatccaggaagaccgagcctacac 3072
Db 2999 caattttagcttcgaatgtagagatatacaggtatccaggtatccaggaagaccgagcctacac 3058
QY 3073 ccgagatccccaaggttcaagaactccccaacaccttcttgagagagccctacacaggg 3132
Db 3059 ccgagatccccaaggttcaagaactccccaacaccttcttgagagagccctacacaggg 3118
QY 3133 accgtgcacactcaagatcccaacacctccaggtatgagaggtatccaggaagcactactgcctg 3192
Db 3119 accgtgcacactcaagatcccaacacctccaggtatgagaggtatccaggaagcactactgcctg 3178
QY 3193 tgccttgcgagagccacaacacagagctgttgaagaaggtatccaggaagcactactgcctg 3252
Db 3179 tgccttgcgagagccacaacacagagctgttgaagaaggtatccaggaagcactactgcctg 3238
QY 3253 aatgtctgaactgaaggtatcaagacctgtctgaagaaggtatccaggaagcactactgcctg 3312
Db 3239 aatgtctgaactgaaggtatcaagacctgtctgaagaaggtatccaggaagcactactgcctg 3298
QY 3313 taacaacttgggtatcaggttgcgagcggcagcagatgctgaagcagagccacaggaaga 3372
Db 3299 taacaacttgggtatcaggttgcgagcggcagcagatgctgaagcagagccacaggaaga 3358
QY 3373 aaactgtatccagatccagccccaacacagcagcacaagtaagagatcttgggga 3432
Db 3359 aaactgtatccagatccagccccaacacagcagcacaagtaagagatcttgggga 3418
QY 3433 cagctggaatttgcagagatgtagatcccggttgcagaccttaagcagcccaactac 3492
Db 3419 cagctggaatttgcagagatgtagatcccggttgcagaccttaagcagcccaactac 3478

QY 3493 cgtacaacaaagaaagggggtatctccctggtctctgaagcacaagagcatttgatg 3552
Db 3479 cgtacaacaaagaaagggggtatctccctggtctctgaagcacaagagcatttgatg 3538
QY 3553 ctatcaaaaagccctgtctgtagcgaactgtctctggtccctccctgtagcgaactaaacct 3612
Db 3539 ctatcaaaaagccctgtctgtagcgaactgtctctggtccctccctgtagcgaactaaacct 3598
QY 3613 ttaccctttagtgatgaagcgttaagggagtagcccgagaggttttaaccacaacctag 3672
Db 3599 ttaccctttagtgatgaagcgttaagggagtagcccgagaggttttaaccacaacctag 3658
QY 3673 gaacatgagagagacactgtgtgtctacactgttcaaaagaagttatccctgtagcagtggt 3732
Db 3659 gaacatgagagagacactgtgtgtgtctacactgttcaaaagaagcttgatccctgtagcagtggt 3718
QY 3733 ggcgcgtatgtctgaaggtctatccgagcgtgtgtgcatatgtgtcaagagcctgaacat 3792
Db 3719 ggcgcgtatgtctgaaggtctatccgagcgtgtgtgcatatgtgtcaagagcctgaacat 3778
QY 3793 tgaacttggagacaagataaactgttaatagcccccatgtcatgtgagacaatcgtctggtc 3852
Db 3779 tgaacttggagacaagataaactgttaatagcccccatgtcatgtgagacaatcgtctggtc 3838
QY 3853 agcccccaagacccgatgtgtagacaaagcccgcatgacacatatacaagcctgtcttca 3912
Db 3839 agcccccaagacccgatgtgtagacaaagcccgcatgacacatatacaagcctgtcttca 3898
QY 3913 cagaagaggtatcaacttgcctccacagcgtcttcaacacctgtcacactctgtctgaag 3972
Db 3899 cagaagaggtatcaacttgcctccacagcgtcttcaacacctgtcacactctgtctgaag 3958
QY 3973 agactgtatgaacccagtgatcatatgtatgtccatcaacattgattgtagagagaggtgtcc 4032
Db 3959 agactgtatgaacccagtgatcatatgtatgtccatcaacattgattgtagagagaggtgtcc 4018
QY 4033 gcaagagacttcaacagataacacgtgtgtagagaaaggttaacattgttcaactgaacgaa 4092
Db 4019 gcaagagacttcaacagataacacgtgtgtagagaaaggttaacattgttcaactgaacgaa 4078
QY 4093 gcaactctgtgtgtgaaggttaagagatggtgtggtggtggtggtggtggtggtggtggt 4152
Db 4079 gcaactctgtgtgtgaaggttaagagatggtgtggtggtggtggtggtggtggtggtggt 4138
QY 4153 cgaactgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4212
Db 4139 cgaactgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4198
QY 4213 tcaagcaagcttgcgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4272
Db 4199 tcaagcaagcttgcgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4258
QY 4273 atgcttgcgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4332
Db 4259 atgcttgcgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4318
QY 4333 cagcaggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4392
Db 4319 cagcaggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4378
QY 4393 tgcataaaggttaacttatatacgttccctgtgagatcaagaagccaaagatctat 4452
Db 4379 tgcataaaggttaacttatatacgttccctgtgagatcaagaagccaaagatctat 4438
QY 4453 cttaggggaacagatgtgtgtagcgtgttgcagagcagcagccagcgtgttgaacctc 4512
Db 4439 cttaggggaacagatgtgtgtagcgtgttgcagagcagcagccagcgtgttgaacctc 4498
QY 4513 tgcctataataagaacgcccagaagccccaagacagagcagagcagcagcagcagcagcagc 4572
Db 4499 tgcctataataagaacgcccagaagccccaagacagagcagagcagcagcagcagcagcagc 4558

QY 4573 ggcagaagataaaagatagaccagttctctggaactccggaggagacccgtatatact 4632
|||||
Db 4559 ggcagaagataaaagataagaccagttctctggaactccggaggagacccgtatatact 4618
QY 4633 catatgggaaggaaatcccgccccacaagaagggttagaatagttccaaacagataatc 4632
|||||
Db 4619 catatgggaaggaaatcccgccccacaagaagggttagaatagttccaaacagataatc 4678
QY 4693 gtctaacccaactaggaactaaacaccgagcaggttggtcagaatcccttatactg 4752
|||||
Db 4679 gtctaacccaactaggaactaaacaccgagcaggttggtcagaatcccttatactg 4738
QY 4753 ttctggaagctaacggaggtggtcgactcgggttggttcaaatgtgtgcccgtgacagctg 4812
|||||
Db 4739 ttctggaagctaacggaggtggtcgactcgggttggttcaaatgtgtgcccgtgacagctg 4798
QY 4813 ttaatgtctaactcttccagaataactccagaagaagactaaagggaagccaccagagcg 4872
|||||
Db 4799 ttaatgtctaactcttccagaataactccagaagaagactaaagggaagccaccagagcg 4858
QY 4873 ctcaactgggaagtggaacttcaacttgaaggtaagccgggttaaatataggaataatatactat 4932
|||||
Db 4859 ctcaactgggaagtggaacttcaacttgaaggtaagccgggttaaatataggaataatatactat 4918
QY 4933 tgggttttgtagacaacttctcaaggatgggttagaggttacttactaagaagaagact 4992
|||||
Db 4919 tgggttttgtagacaacttctcaaggatgggttagaggttacttactaagaagaagact 4978
QY 4993 caaccgtgtgtgcttaagaataatctcggaggaatcttccaaagattggagataactaag 5052
|||||
Db 4979 caaccgtgtgtgcttaagaataatctcggaggaatcttccaaagattggagataactaag 5038
QY 5053 taataggtgcagacaatactgtccagcttcggttgcagagtaagtcagggagcttgccaga 5112
|||||
Db 5039 taataggtgcagacaatactgtccagcttcggttgcagagtaagtcagggagcttgccaga 5098
QY 5113 tatggggatgtgattgnaaactgtcatgtgtcataagaccccaagctccagacaggtag 5172
|||||
Db 5099 tatggggatgtgattgnaaactgtcatgtgtcataagaccccaagctccagacaggtag 5158
QY 5173 aagggatggaataagaaaccttaagaagaccccttaaccaaatgtgacacagagactgcat 5232
|||||
Db 5159 aagggatggaataagaaaccttaagaagaccccttaaccaaatgtgacacagagactgcat 5218
QY 5233 atgattggaatgtctctcgtcccttggtgtctttagaggtagagaaaccccttgacagt 5292
|||||
Db 5219 atgattggaatgtctctcgtcccttggtgtctttagaggtagagaaaccccttgacagt 5278
QY 5293 ttgggtcgtgaaccccttataatgtgtctctacgggggaccccccgttggcagaatgtcct 5352
|||||
Db 5279 ttgggtcgtgaaccccttataatgtgtctctacgggggaccccccgttggcagaatgtcct 5338
QY 5353 ttgacatagtgtgattgtgattgtcttccaaacttggtctctataggtccaaagggtcgtg 5412
|||||
Db 5339 ttgacatagtgtgattgtgattgtcttccaaacttggtctctataggtccaaagggtcgtg 5398
QY 5413 agtgggtgagagcagagcgttggaagagagcttcgggagagctctaaaggaagagacttgc 5472
|||||
Db 5399 agtgggtgagagcagagcgttggaagagagcttcgggagagctctaaaggaagagacttgc 5458
QY 5473 aagttccacatcgcttccaaagttagagattcaagttcattgtataagcccaacggtgcagaa 5518
|||||
Db 5459 aagttccacatcgcttccaaagttagagattcaagttcattgtataagcccaacggtgcagaa 5518
QY 5533 accctggaactcgtgtggaagagacttatactcgtacttctgacacacacacagcgctgtga 5592
|||||
Db 5519 accctggaactcgtgtggaagagacttatactcgtacttctgacacacacacagcgctgtga 5578
QY 5593 aagtcgaagaaatcccccacttgatccaatgatacccaagcttaagccggcgccacttccg 5652
|||||
Db 5579 aagtcgaagaaatcccccacttgatccaatgatacccaagcttaagcttccgacccctccg 5638
QY 5653 actcgggtgtgagagccgaagaagactggaatcccttaagcttcgctccatcgctggtg 5712
|||||
Db 5639 acccgggtgtgagagccgaagaagactggaatcccttaagcttcgctccatcgctggtg 5698
QY 5713 ttcttactactaaacataactccccaaggacatgaataagccttataagacgtcgacac 5772
|||||
Db 5699 ttcttactactaaacataactccccaaggacatgaataagccttataagacgtcgacac 5758
QY 5773 ccccatagacacttatacccttaactcgtcgtatattgaacccctgataaggggtgtaactgta 5832
|||||
Db 5759 ccccatagacacttatacccttaactcgtcgtatattgaacccctgataaggggtgtaactgta 5818
QY 5833 aatagcactcgaaggtgtgtctcctagaagaccccttggtgagcctgaactgattctgcctc 5892
|||||
Db 5819 aatagcactcgaaggtgtgtctcctagaagaccccttggtgagcctgaactgattctgcctc 5878
QY 5893 cgtattgttaaccccgctgtttaaagacacacccccaactagtcggtataggttc 5952
|||||
Db 5879 cgtattgttaaccccgctgtttaaagacacacccccaactagtcggtataggttc 5938
QY 5953 tattgtgtcccaaggcacagagaagaagaatactgtggtggttctgtgggaatcctcgt 6012
|||||
Db 5939 tattgtgtcccaaggcacagagaagaagaatactgtggtggttctgtgggaatcctcgt 5998
QY 6013 agggatgagagctcgttcaactcccaagatggagagactgggaatggcgcgactctccag 6072
|||||
Db 5999 agggatgagagctcgttcaactcccaagatggagagactgggaatggcgcgactctccag 6058
QY 6073 gacccgggttaaaattctccttctgaattccggccgggcaagtlacaaatgtatgaacta 6132
|||||
Db 6059 gacccgggttaaaattctccttctgaattccggccgggcaagtlacaaatgtatgaacta 6118
QY 6133 tataagaataagagctgtcctccatcaagacttgaattatataagaataagtttcaactga 6192
|||||
Db 6119 tataagaataagagctgtcctccatcaagacttgaattatataagaataagtttcaactga 6178
QY 6193 aggaataacaggaataatatacctaaagtggaataatgtatgagccgggtgaatagtttcta 6252
|||||
Db 6179 aggaataacaggaataatatacctaaagtggaataatgtatgagccgggtgaatagtttcta 6238
QY 6253 ttatagcgggggagcaggttccacttaacacttggtccttaagataagacgggagca 6312
|||||
Db 6239 ttatagcgggggagcaggttccacttaacacttggtccttaagataagacgggagca 6298
QY 6313 gaaccccggttggcgaatggagccgaataaagtactgtggaacagggcccgccgtg 6372
|||||
Db 6299 gaaccccggttggcgaatggagccgaataaagtactgtggaacagggcccgccgtg 6358
QY 6373 gaaccccggttggcgaatggagccgaataaagtactgtggaacagggcccgccgtg 6332
|||||
Db 6359 gaaccccggttggcgaatggagccgaataaagtactgtggaacagggcccgccgtg 6318
QY 6433 ccgctctgacaaagttccacttgatattcttcaacacacacgcttagaactcccaagt 6492
|||||
Db 6419 ccgctctgacaaagttccacttgatattcttcaacacacacgcttagaactcccaagt 6478
QY 6493 gttccttgaagacagagacagacttcaagtctcatcccaagggacttccaagcctac 6552
|||||
Db 6479 gttccttgaagacagagacagacttcaagtctcatcccaagggacttccaagcctac 6538
QY 6553 aactccacgaaccccttgatgcacactctctgtgtgtcgttctatccctgaagcctcct 6612
|||||
Db 6539 aactccacgaaccccttgatgcacactctctgtgtgtcgttctatccctgaagcctcct 6598
QY 6613 tattaatgagggatgtgcttaagaagaagaataatcaatgtgacaaagagacataagaatcca 6672
|||||
Db 6599 tattaatgagggatgtgcttaagaagaagaataatcaatgtgacaaagagacataagaatcca 6658
QY 6673 tttacatgggggttcccgaaataaagcttaacccctgaaggttcccggaaggagacatgc 6732
|||||
Db 6659 tttacatgggggttcccgaaataaagcttaacccctgaaggttcccggaaggagacatgc 6718
QY 6733 ataggaagaagctccccaatcccaacacacttgtatagactgtgtgttatggcag 6792
|||||

us-09-171-553b-3.feb2201i.rng

xx Sequence 7333 BP; 1984 A; 1788 C; 1885 G; 1676 T; 0 other;
sq

36.1%; Score 2967; DB 22; Length 7333;

Query Match	36.16	Score 2507	25
Best Local Similarity	99.8%	Pred. No. 0;	
Matches 3507;	Conservative	0; Mismatches	6; Indels 2; Gaps 2

QY	1	gtgtgtgaacacgtgtggtcccaagtcgctcgtggaataaaatccctcgtgttgatc	70
Db	11	gtgtgtgaacacgtgtggtcccaagtcgctcgtggaataaaatccctcgtgttgatc	70
QY	61	aagacgcctctcgttaagtgatctgggtgtgtccctctccagaccggaacgaagggat	120
Db	71	aagacgcctctcgttaagtgatctgggtgtgtccctctccagaccggaacgaagggat	130
QY	121	tgctcttacttgaccttcaattgctgtgtcgccggaataccctcggaaccaacctta	180
Db	131	tgctcttacttgaccttcaattgctgtgtcgccggaataccctcggaaccaacctta	190
QY	181	caaccgagaacccgacttggaaagttaaagggtaccctctggaagctgtgtgtgtgcgc	240
Db	191	caaccgagaacccgacttggaaagttaaagggtaccctctggaataatgtgtgtgcgc	250
QY	241	ggcgcctcgtctcgtatgtctgtcttccgtgtatgtcgccgtcttcggtttgcacatctcc	300
Db	251	ggcgcctcgtctcgtatgtctgtcttccgtgtatgtcgccgtcttcggtttgcacatctcc	310
QY	301	ctcaagacgttaagacatgtagagacgtgtatacgaagaagtgctcgtgagagatcacagctg	360
Db	311	ctcaagacgttaagacatgtagagacgtgtatacgaagaagtgctgagagatcacagctg	370
QY	361	ccacccttggtggaacgcgcccggtgaggttggtggaagaccgaagacgcctgtgtctccact	420
Db	371	ccacccttggtggaacgcgcccggtgaggttggtggaagaccgaagacgcctgtgtctccact	430
QY	421	gtcgtgtcaagagacgcgactctcgtgtgtgtgaagcgaacacctccacctccgcgcgtccg	480
Db	431	gtcgtgtcaagagacgcgactctcgtgtgtgtgaagcgaacacctccacctccgcgcgtccg	490
QY	481	acctctttgtcctcgtgtgtggaagacgcgcgcgcgcgcgtcgtgtctgtgatctgtgtgtt	540
Db	491	acctctttgtcctcgtgtgtggaagacgcgcgcgcgcgcgtcgtgtctgtgatctgtgtgtt	550
QY	541	ctgcctcgtgtgtcttctgtctgtctgtccctgtgtctcaagctttaaataatggaacgaag	600
Db	551	ctgtctcgtgtgtcttctgtctgtctgtccctgtgtctcaagctttaaataatggaacgaag	610
QY	601	tgactaaccccccttaagttgactctcgcgaacattggaactgaagttagaatccagggctcata	660
Db	611	tgactaaccccccttaagttgactctcgcgaacattggaactgaagttagaatccagggctcata	670
QY	661	atttgatcagttcaagttaaagaaggaccttggcagaagactctctgtcctctgaaatgcca	720
Db	671	atttgatcagttcaagttaaagaaggaccttggcagaagactctctgtcctctgaaatgcca	730
QY	721	cattcgaatgttgaaatgggcacacacagagggaaccttaattctgaaatataccctgtgcgtta	780
Db	731	cattcgaatgttgaaatgggcacacacagagggaaccttaattctgaaatataccctgtgcgtta	790
QY	781	aagcaatcatttttcaagatcgtgaccgcgcctctatccctgatacagaagccctatactcta	840
Db	791	aagcaatcatttttcaagatcgtgaccgcgcctctatccctgatacagaagccctatactcta	850
QY	841	cgtgtgcaaatgttgcaagaagatctctccgcacatgggtttaaacacatggtcgtaaataaccaa	900
Db	851	cgtgtgcaaatgttgcaagaagatctctccgcacatgggtttaaacacatggtcgtaaataaccaa	910
QY	901	gaagacgaagtcccgaaatccctgtgctcttggtggaagaaaaaacaacatccgcgcgaataag	960
Db	911	gaagacgaagtcccgaaatccctgtgctcttggtggaagaaaaaacaacatccgcgcgaataag	970
QY	961	tcgagccctctctcgtatctaccccggaatcgaagacgcgcgcgcgcgcgcgcgcgcgcgcgc	1020

Db	971	tcgagccctc-tcctcgtaactaccccgagatcagagagccgacacttgcgcggaaacc	1029
Qy	1021	caaccctgtccccaacccttaccagacagagtgctgtgaggggaa-cctctgcccct	1079
Db	1030	caactgtctccccaacccttaccagacagagtgctgtgaggggaaaccctctgcccct	1089
Qy	1080	cctgtagctccggtgtgtgagggagacctgcgtccgaggaactcggagaccggagagcgcaac	1139
Db	1090	cctgtagctccggtgtgtgagggagacctgcgtccgaggaactcggagagcgagagcgcaac	1149
Qy	1140	ccggagcggagacagagagatcggagataaccgctgtgcacactatgagccctccatgcca	1199
Db	1150	ccggagcggagacagagagatcggagataaccgctgtgcacactatgagccctccatgcca	1209
Qy	1200	ggggggccaattgcagccccctcaatltgaccttctctctgcagatactctataattg	1259
Db	1210	ggggggccaattgcagccccctcaatltgaccttctctctgcagatactctataattg	1269
Qy	1260	aaactaaacatccccccttctcggagagatccccaagccctcaaggggttgggtggagctcc	1319
Db	1270	aaactaaacatccccccttctcggagagatccccaagccctcaaggggttgggtggagctcc	1329
Qy	1320	cttatgtctctcaacagactactctggagatgattgtcaacagctgtgtcgacactcttc	1379
Db	1330	cttatgtctctcaacagactactctggagatgattgtcaacagctgtgtcgacactcttc	1389
Qy	1380	acacacgagagcagagagagaattctgttagaggtctagaaaaaagtctctcgtgggcccagac	1439
Db	1390	acacacgagagcagagagagaattctgttagaggtctagaaaaaagtctctcgtgggcccagac	1449
Qy	1440	gggcgaccccaagctgttgcaaaatgagatcttgacatggaattctcccttgactgcgcgggt	1499
Db	1450	gggcgaccccaagctgttgcaaaatgagatcttgacatggaattctcccttgactgcgcgcgt	1509
Qy	1500	tggagctacacaacacgctgtgaaggttagggagagcttgaaaatctatccgacagctctgtg	1559
Db	1510	tggagctacacaacacgctgtgaaggttagggagagcttgaaaatctatccgacagctctgtg	1569
Qy	1560	ggcggtctccgggctgcctcaagaagccggccacacttctgtcctaaggttaaagagagtgtatg	1619
Db	1570	ggcggtctccgggctgcctcaagaagccggccacacttctgtcctaaggttaaagagagtgtatg	1629
Qy	1620	caaggagacagaaacacccctccctgtattctcttgagaggtctcatgtgaagccttcaaggcg	1679
Db	1630	caaggagacagaaacacccctccctgtattctcttgagaggtctcatgtgaagccttcaaggcg	1689
Qy	1680	ttcacaccctttagtactactcctcagagggcccaagaagcctcagatgtgcgcctgcttcat	1739
Db	1690	ttcacaccctttagtactactcctcagagggcccaagaagcctcagatgtgcgcctgcttcat	1749
Qy	1740	gggcagtcgggtcttgatatacagagaagaaattcagagagcttggaaggtttacagggagct	1799
Db	1750	gggcagtcgggtcttgatatacagagaagaaattcagagagcttggaaggtttacagggagct	1809
Qy	1800	gaattacgtgatctagtgagagagcgagagaaagtgttatccagaagggagaaacagaaagag	1859
Db	1810	gaattacgtgatctagtgagagagcgagagaaagtgttatccagaagggagaaacagaaagag	1869
Qy	1860	gagagaggaacagagaaagaaagagagagagaaagggagagagagcggtgatatacgg	1919
Db	1870	gagagaggaacagagaaagaaagagagagagaaagggagagagagcggtgatatacgg	1929
Qy	1920	caagagagaagatttgactaagatctgtgcgcgagtggttgaagggaaagagcagcagggag	1979
Db	1930	caagagagaagatttgactaagatctgtgcgcgagtggttgaagggaaagagcagcagggag	1989
Qy	1980	agagagagagattttagaaaaattaggttcagggcctcagacagctcagagaccttggcaat	2039
Db	1990	agagagagagatttttagaaaaattaggttcagggcctcagacagctcagagaccttggcaat	2049
Qy	2040	aggagcccaatccgacagagccagtgctgtgatttgaagaaaaaggaacatgggccaag	2099

```

Db 2050 aggagcccaactcgcacaaaggaccagtgctgattgttaagaaaaaggacactcggcgaag 2109
QY 2100 aactgcccaagaaggggaaacaaaggccgaaggtcctaactctagaagaataaagat 2159
Db 2110 aactgcccaagaaggggaaacaaaggccgaaggtcctaactctagaagaataaagat 2169
QY 2160 taggggaacgggtgttcggaccccccccgagcccgaggttaactcttgaagggtgaggg 2219
Db 2170 taggggaacgggtgttcggaccccccccgagcccgaggttaactcttgaagggtgaggg 2229
QY 2220 caacgaactgagttcctgtgttaacgggagcgagcatcagctgtctacaacatla 2279
Db 2230 caacgaactgagttcctgtgttaacgggagcgagcatcagctgtctacaacatla 2289
QY 2280 ggaataactaaagaaaaaaatccctgggtgatgggtggtccacgggcaacggcagatcca 2339
Db 2290 ggaataactaaagaaaaaaatccctgggtgatgggtggtccacgggcaacggcagatcca 2349
QY 2340 tggactacccgaagaacccgttgactgtggaggtgggagcggtlaacccactcgtctcgtc 2399
Db 2350 tggactacccgaagaacccgttgactgtggaggtgggagcggtlaacccactcgtctcgtc 2409
QY 2400 atccctgagtgccagttaccctcttaggtagaagacttaactgacccaagatgggagctcaa 2459
Db 2410 atccctgagtgccagttaccctcttaggtagaagacttaactgacccaagatgggagctcaa 2469
QY 2460 attcttcttgaacaaagggaagcagaagtgctgtgaatacaaaacccactcgtgtg 2519
Db 2470 attcttcttgaacaaagggaagcagaagtgctgtgaatacaaaacccactcgtgtg 2529
QY 2520 accctccaattagatgatgatatacgaatactctcccaaglaaagcctgtgtaagat 2579
Db 2530 accctccaattagatgatgatatacgaatactctcccaaglaaagcctgtgtaagat 2589
QY 2580 atacagtcctgtgtggaacagttccccaagcctgggacagaacccgacaggaatgggtg 2639
Db 2590 atacagtcctgtgtggaacagttccccaagcctgggacagaacccgacaggaatgggtg 2649
QY 2640 gcaaggaagttccccaacaggtatcaactgaagcaggtgttaccacatcaagtc 2699
Db 2650 gcaaggaagttccccaacaggtatcaactgaagcaggtgttaccacatcaagtc 2709
QY 2700 agacagtaacccctgagtagaaggtcgcgaagaagatttggcgcagatgttcaaaagata 2759
Db 2710 agacagtaacccctgagtagaaggtcgcgaagaagatttggcgcagatgttcaaaagata 2769
QY 2760 atccaacagggcaltcctgtcctgtcccaatccctctggaatacctccctgtcaccggt 2819
Db 2770 atccaacagggcaltcctgtcctgtcccaatccctctggaatacctccctgtcaccggt 2829
QY 2820 aggaagcctgggacccatattatcgaacgtacagatcgaagatgagaggtcaataaaagg 2879
Db 2830 aggaagcctgggacccatattatcgaacgtacagatcgaagatgagaggtcaataaaagg 2889
QY 2880 gtgcaggaatacaacccaacggtcccgaaaccttataacccctgtgagcgccctccgct 2939
Db 2890 gtgcaggaatacaacccaacggtcccgaaaccttataacccctgtgagcgccctccgct 2949
QY 2940 gaacggaactgttacaacgtatgtgacttaaaagatgctctctcctcgagataacag 2999
Db 2950 gaacggaactgttacaacgtatgtgacttaaaagatgctctctcctcgagataacag 3009
QY 3000 cccactagccaacacacttctgcttgcgaatggagagatcccaagtcagggaacccgg 3059
Db 3010 cccactagccaacacacttctgcttgcgaatggagagatcccaagtcagggaacccgg 3069
QY 3060 cagctcacttgagcccgactgcgccaaaggttcaagaactccocgaacactcttgagaa 3119
Db 3070 cagctcacttgagcccgactgcgccaaaggttcaagaactccocgaacactcttgagaa 3129
QY 3120 gccctacacaggaagcctgcgcacacttcaagatccaacccctcaagtgacccctccag 3179
Db 3130 gccctacacaggaagcctgcgcacacttcaagatccaacccctcaagtgacccctccag 3189

```

```

QY 3180 tacgtgatgacctgtcttcggcgagagcccaacagagactgttagaaggtacgaag 3239
Db 3190 tacgtgatgacctgtcttcggcgagagcccaacagagactgttagaaggtacgaag 3249
QY 3240 gcaactcgtctggaattcttgacctagaggtcacaagagcctcgtctaagaagccagat 3299
Db 3250 gcaactcgtctggaattcttgacctagaggtcacaagagcctcgtctaagaagccagat 3309
QY 3300 tgcagagagaggttaacatacttgggttacaagtttgcggggcgagcagatgagcgaag 3359
Db 3310 tgcagagagaggttaacatacttgggttacaagtttgcggggcgagcagatgagcgaag 3369
QY 3360 gagcagcggaagaacactgtagtcacagatccggcccaacacagccaagaatgaga 3419
Db 3370 gagcagcggaagaacactgtagtcacagatccggcccaacacagccaagaatgaga 3429
QY 3420 gagtttctgggagacagctggaatttgcagactgtgatacccggtttgagacctaga 3479
Db 3430 gagtttctgggagacagctggaatttgcagactgtgatacccggtttgagacctaga 3489
QY 3480 gccccactctacccgcgttaacccaagaagaagggg 3514
Db 3490 gccccactctacccgcgttaacccaagaagaagggg 3524

```

RESULT 4
AAT74883 standard; cDNA; 7393 BP.

ID AAT74883;
AC AAT74883;
XX 09-FEB-1998 (first entry)

DT Porcine retrovirus cDNA (defective).
DE
XX Porcine retrovirus. GAG protein; ENV protein;
KW xenotransplantation; infectious; provirus; organ transplant; donor;
KW activated virus; PCR; ss.
XX OS Porcine retrovirus.

XX FH Key Location/Qualifiers
FH CDS 598..2172
FT /*tag= a
598..2169
FT /*tag= b
FT /*note= "putative GAG protein"
FT CDS 2320..4737
FT /*tag= c
FT /*note= "putative POL coding region (partial) as
described in the specification"
FT mat_peptide 2320..3522
FT /*tag= d
FT /*note= "putative POL protein (partial)"
FT mat_peptide 3516..4328
FT /*tag= e
FT /*note= "putative POL protein (partial)"
FT CDS 4332..4748
FT /*tag= f
FT 4738..6725
FT /*tag= g
FT /*note= "putative ENV coding region (partial) as
described in the specification"
FT mat_peptide 4752..6722
FT /*tag= h
FT /*note= "ENV protein (partial)"
XX MO9721036-A1.
XX 19-JUN-1997.
XX 13-DEC-1996; 96WO-US19680.

XX 14-DEC-1995; 9505-0572645.
 XX (GENO) GEN HOSPITAL CORP.
 XX Fishman JA:
 XX WPI: 1997-332804/30.
 XX P-PSDB; AAW32091-W32095.
 XX New nucleic acid from porcine retroviruses - used for detecting
 XX viruses in transplant or other tissue and for assessing risk of
 XX transmitting infection to graft recipient
 XX Claim 16; Fig 2; 128pp; English.
 XX This CDNA sequence represents a defective purified swine retrovirus
 XX found in PK-15 cells containing the putative coding regions for viral
 XX GAG, POL and ENV proteins. There are a few in frame stop codons and
 XX apparent frame shifts in the given coding sequence which alter features
 XX of the translation. This sequence and PCR fragments generated from the
 XX of the translation. This sequence and PCR fragments generated from the
 XX presence (see AAT74812-T74882) could be used to screen organs for the
 XX presence of porcine retroviruses prior to xenotransplantation.
 XX Transplantation can increase the likelihood of retroviral activation if
 XX intact and infectious proviruses are present. The porcine retroviral
 XX sequence can be used to generate probes to determine the level (e.g.
 XX copy number) of intact (i.e. potentially replicating) porcine provirus
 XX sequences in a strain of xenograft transplantation donors. It can be
 XX used to detect mutations, genetic lesions or viral recombinants and to
 XX determine the histological localisation of activated retroviruses. Using
 XX Polymerase Chain Reaction DNA Quantitation (PQ) on blood mononuclear
 XX cells, infectivity titration and susceptibility testing can be
 XX performed. Ultimately animal donors without intact porcine retroviral
 XX sequences or a lower copy number of viral elements could be selected.

XX Sequence 7393 BP; 2004 A; 1801 C; 1896 G; 1692 T; 0 other;

Query Match 36.1%; Score 2967; DB 18; Length 7393;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 1 ggtgtgtacagactgtggccagcgctgtgataaataaccccttgcattgacac 60
 Db 11 gttgtgtacagactgtggccagcgctgtgataaataaccccttgcattgacac 70
 QY 61 aagacgctctctgtgaattgtgggtgtcgcctcttcgcagccggaagagggat 120
 Db 71 aagacgctctctgtgaattgtgggtgtcgcctcttcgcagccggaagagggat 130
 QY 121 tcttcttactgaccttcaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
 Db 131 tcttcttactgaccttcaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 190
 QY 181 caaccgagagacgacttgagaaagaaaccccttggaaacgtgtgtgtgtgtgtgt 240
 Db 191 caaccgagagacgacttgagaaagaaaccccttggaaacgtgtgtgtgtgtgtgt 250
 QY 241 ggcgtctgt 300
 Db 251 ggcgtctgt 310
 QY 301 ctacagacgagagacttgagaaagaaaccccttggaaacgtgtgtgtgtgtgtgt 360
 Db 311 ctacagacgagagacttgagaaagaaaccccttggaaacgtgtgtgtgtgtgtgt 370
 QY 361 caaccctgt 420
 Db 371 caaccctgt 430
 QY 421 gtcgt 480
 Db 431 gtcgt 490

QY 481 actcttgcctgtctgtgtgaagacgcgagcggtgcgtgtgtgtgtgtgtgtgtgt 540
 Db 491 actcttgcctgtctgtgtgaagacgcgagcggtgcgtgtgtgtgtgtgtgtgtgt 550
 QY 541 ctgtctgt 600
 Db 551 ctgtctgt 610
 QY 601 tgaactcccccttaattgtgactctgcacattggagctaaagttaacagagctata 660
 Db 611 tgaactcccccttaattgtgactctgcacattggagctaaagttaacagagctata 670
 QY 661 attgtcagttcagtttaagaagagacttgagacattctgtgccttgaaatggcga 720
 Db 671 attgtcagttcagtttaagaagagacttgagacattctgtgccttgaaatggcga 730
 QY 721 catcagattgtgagtcacatcaagaagagacatttcaattcgaattatcctgctgt 780
 Db 731 catcagattgtgagtcacatcaagaagagacatttcaattcgaattatcctgctgt 790
 QY 781 aggaacatcttttcagacttgagacccgctcctcattcagatcagagccctataccta 840
 Db 791 aggaacatcttttcagacttgagacccgctcctcattcagatcagagccctataccta 850
 QY 841 ctgtgtcagaattgtgcagaagatctcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
 Db 851 ctgtgtcagaattgtgcagaagatctcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 910
 QY 901 gaaagcaggtctccggaatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960
 Db 911 gaaagcaggtctccggaatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 970
 QY 961 tgcagccctctcctgtatctaccccgagatcagagagccgcgcgcgcgcgcgcgcgc 1020
 Db 971 tgcagccctctcctgtatctaccccgagatcagagagccgcgcgcgcgcgcgcgcgc 1029
 QY 1021 caactgttcccccaccccttaccagacaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1079
 Db 1030 caactgttcccccaccccttaccagacaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1089
 QY 1080 cctgtgagctcgcgt 1139
 Db 1090 cctgtgagctcgcgt 1149
 QY 1140 cctgtgagctcgcgt 1199
 Db 1150 cctgtgagctcgcgt 1209
 QY 1200 gggagccaatltgagccctccagatattgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1259
 Db 1210 gggagccaatltgagccctccagatattgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1269
 QY 1260 aaacttaacatccctcttctcgcagagatccccaagcctcgaaggtgtgtgtgtgt 1319
 Db 1270 aaacttaacatccctcttctcgcagagatccccaagcctcgaaggtgtgtgtgtgt 1329
 QY 1320 ctatgttcttccacagcctacttgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1379
 Db 1330 ctatgttcttccacagcctacttgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1389
 QY 1380 acaacgagagagcagagagaaatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1439
 Db 1390 acaacgagagagcagagagaaatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1449
 QY 1440 gggcgaacacagcagctgtgcaaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1499
 Db 1450 gggcgaacacagcagctgtgcaaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1509
 QY 1500 tggagactcaacaacagcgtgtgaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1559
 Db 1510 tggagactcaacaacagcgtgtgaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1569

QY 1560 ggggtctcgggctcgggctcgaagacggcccaatttgctgaagtaagagagtgatg 1619
 Db 1570 ggggtctcgggctcgggctcgaagacggcccaatttgctgaagtaagagagtgatg 1629
 QY 1620 cagggaaccgaagaaacccctctggtatctcttgagagagctcagaaagccttcagagc 1679
 Db 1630 cagggaaccgaagaaacccctctggtatctcttgagagagctcagaaagccttcagagc 1689
 QY 1680 ttcccccttttgatctactacccagaagggcccaagagctcaggtccctgctccat 1739
 Db 1690 ttcccccttttgatctactacccagaagggcccaagagctcaggtccctgctccat 1749
 QY 1740 gggagctcggctcggatcaggaagaaacttcagagactcagagaggttcaagagagct 1799
 Db 1750 gggagctcggctcggatcaggaagaaacttcagagactcagagaggttcaagagagct 1809
 QY 1800 gagttacgtgatctcagtgagagagcagagagagtgatcttcagaaagggagacagagag 1859
 Db 1810 gagttacgtgatctcagtgagagagcagagagagtgatcttcagaaagggagacagagag 1869
 QY 1860 gagaaagaaacagaaagaaagaaagagagagagagagagagagagagagagagagag 1919
 Db 1870 gagaaagaaacagaaagaaagaaagagagagagagagagagagagagagagagagag 1929
 QY 1920 cagagaaagaaacttgactaagatctcggcagtggttgtaagggagagagagagagag 1979
 Db 1930 cagagaaagaaacttgactaagatctcggcagtggttgtaagggagagagagagagag 1989
 QY 1980 aagagagagagatcttgagaaataatagtgacagggccctcagagacagagagagagag 2039
 Db 1990 aagagagagagatcttgagaaataatagtgacagggccctcagagacagagagagagag 2049
 QY 2040 aggaaccacacacgaacgaacagtggtgctatgttaagaaagaaagacacacagagagag 2099
 Db 2050 aggaaccacacacgaacgaacagtggtgctatgttaagaaagaaagacacacagagagag 2109
 QY 2100 aactgcccagaagagagagaaacaaagagacggaggtcctcagctcagagagagagagag 2159
 Db 2110 aactgcccagaagagagagaaacaaagagacggaggtcctcagctcagagagagagagag 2169
 QY 2160 tagggagagagaggtctcggaccccccctcggagggagagagagagagagagagagag 2219
 Db 2170 tagggagagagaggtctcggaccccccctcggagggagagagagagagagagagagag 2229
 QY 2220 caaccagttgaggtctcgttgatccggagcagagacatcagtggtcgtacaaacacat 2279
 Db 2230 caaccagttgaggtctcgttgatccggagcagagacatcagtggtcgtacaaacacat 2289
 QY 2280 ggaagaaactaaagaaataatcctcgtggtgagtggtcggagagagagagagagagag 2339
 Db 2290 ggaagaaactaaagaaataatcctcgtggtgagtggtcggagagagagagagagagag 2349
 QY 2340 tggactaacccgaagagacgggtgaggttgagtggtgagtggtgagtggtgagtggtgag 2399
 Db 2350 tggactaacccgaagagacgggtgaggttgagtggtgagtggtgagtggtgagtggtgag 2409
 QY 2400 atccctggtggtccagtaacccctctagtgagagacttactcagcagagtggtgagag 2459
 Db 2410 atccctggtggtccagtaacccctctagtgagagacttactcagcagagtggtgagag 2469
 QY 2460 attcttttgagaaag 2519
 Db 2470 attcttttgagaaag 2529
 QY 2520 accctcaatcagatcagatcagatcagatcagatcagatcagatcagatcagatcagat 2579
 Db 2530 accctcaatcagatcagatcagatcagatcagatcagatcagatcagatcagatcagat 2589
 QY 2580 atacagtcctggttgag 2639
 Db 2590 atacagtcctggttgag 2649
 QY 2649 gcaag 2699

Db 2650 gcaag 2709
 QY 2700 agacagtaaccccttgag 2759
 Db 2710 agacagtaaccccttgag 2769
 QY 2760 atcccaag 2819
 Db 2770 atcccaag 2829
 QY 2820 aggaagccctgggagacacagatcagatcagatcagatcagatcagatcagatcagatc 2879
 Db 2830 aggaagccctgggagacacagatcagatcagatcagatcagatcagatcagatcagatc 2889
 QY 2880 gtcag 2939
 Db 2890 gtcag 2949
 QY 2940 gaaacgag 2999
 Db 2950 gaaacgag 3009
 QY 3000 cccactagcccaacacacttttgctcagatgagagagagagagagagagagagagagag 3059
 Db 3010 cccactagcccaacacacttttgctcagatgagagagagagagagagagagagagagag 3069
 QY 3060 cagctacacccctgagacccagctcggcagagagagagagagagagagagagagagagag 3119
 Db 3070 cagctacacccctgagacccagctcggcagagagagagagagagagagagagagagagag 3129
 QY 3120 gccctacacag 3179
 Db 3130 gccctacacag 3189
 QY 3180 taagtgag 3239
 Db 3190 taagtgag 3249
 QY 3240 gcaactacgtcgtgagatctcgtcagagagagagagagagagagagagagagagagagag 3299
 Db 3250 gcaactacgtcgtgagatctcgtcagagagagagagagagagagagagagagagagagag 3309
 QY 3300 tgcag 3359
 Db 3310 tgcag 3369
 QY 3360 gaggcag 3419
 Db 3370 gaggcag 3429
 QY 3420 gagtttttgag 3479
 Db 3430 gagtttttgag 3489
 QY 3480 gcccaactacccgtcag 3514
 Db 3490 gcccaactacccgtcag 3524

RESULT 5
 ID AAV09698 standard; DNA; 3320 BP.
 AC AAV09698;
 XX 19-MAY-1998 (first entry)
 DE Porcine retrovirus pol and env DNA.
 XX Porcine retrovirus; POEV; POL protein; ENV protein; vaccine;
 KW diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
 XX

OS Porcine retrovirus.
 XX Key Location/Qualifiers
 XX CDS 23..2793
 FT /tag= a
 FT /product= pol. protein
 FT /note= "polymerase protein"
 FT 2642..3287
 CDS
 FT /tag= D
 FT /product= ENV protein
 FT /note= "envelope protein"
 XX
 XX M09740167-A1.
 XX 30-OCT-1997.
 XX 18-APR-1997; 97WO-GB01087.
 XX 10-FEB-1997; 97GB-0002668.
 XX 19-APR-1996; 96GB-0008164.
 XX (IMUT-) IMUTRAN LTD.
 XX (OONE-) O-ONE BIOTECH LTD.
 XX Galbraith DN, Haworth C, Lees GM, Smith KT;
 XX WPI: 1997-535851/49.
 XX Polynucleotide encoding porcine retrovirus expression product -
 XX useful to develop products for use in vaccines, diagnosis and
 XX xeno-transplantation
 XX
 PS Claim 4; Fig 1: 69pp; English.
 CC This DNA sequence encodes the porcine retrovirus (PoEV) polymerase (POL)
 CC and envelope (ENV) proteins. These proteins can be used to develop viral
 CC vaccines, antisense nucleic acids, ribozymes and other antiviral agents.
 CC They can also be used in xeno-transplantation technology and as
 CC diagnostic tools.
 XX
 SO Sequence 3320 BP; 922 A; 844 C; 832 G; 722 T; 0 other;
 Query Match 32.7%; Score 2682; DB 18; Length 3320;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 381 ttgcggggcgagcgatgctgagcggagcaggaagaaactgagtcagatacc 440
 QY 3392 ggcaccaacacagccaaacagtgagagagttttgggagacgtgatttgcagact 3451
 Db 441 ggcaccaacacagccaaacagtgagagagttttgggagacgtgatttgcagact 500
 QY 3452 gtgagtcgggggttgcagacttaagcagccactctaacccgttaacaaagaagag 3511
 Db 501 gtgagtcgggggttgcagacttaagcagccactctaacccgttaacaaagaagag 560
 QY 3512 gggatctcctgggtcttcagacacagagagatttgatgcatcaaaagagcctgt 3571
 Db 561 gggatctcctgggtcttcagacacagagagatttgatgcatcaaaagagcctgt 620
 QY 3572 gaggcagcctgtctgtgccccttcagcagtaaaccttaaccccttaagtgtgata 3631
 Db 621 gaggcagcctgtctgtgccccttcagcagtaaaccttaaccccttaagtgtgata 680
 QY 3632 ggttaaggagatagccgagagagtttaaccacaccccttaggacatggagagacctgt 3691
 Db 681 ggttaaggagatagccgagagagtttaaccacaccccttaggacatggagagacctgt 740
 QY 3692 tgcctactctgcaagaagagcttgatcctgtagcagatgtgtgcccgtatgttgaaagc 3751
 Db 741 tgcctactctgcaagaagagcttgatcctgtagcagatgtgtgcccgtatgttgaaagc 800
 QY 3752 tatcgagctgtgcccactgtgtaagagcgtgcaacatgtactttggagagagat 3811
 Db 801 tatcgagctgtgcccactgtgtaagagcgtgcaacatgtactttggagagagat 860
 QY 3812 aactgttaatagccccccttgatgtgagagacatgcttgagagccccaagcagatgat 3871
 Db 861 aactgttaatagccccccttgatgtgagagacatgcttgagagccccaagcagatgat 920
 QY 3872 gaccacgcccagatgacccactatcaaaagcgtcttcacagagagaggttcatttcgc 3931
 Db 921 gaccacgcccagatgacccactatcaaaagcgtcttcacagagagaggttcatttcgc 980
 QY 3932 tccacacagcgtcttcaaccccgccactcttgcttaagagagcttgacacagatgac 3991
 Db 981 tccacacagcgtcttcaaccccgccactcttgcttaagagagcttgacacagatgac 1040
 QY 3992 tcatgattgcatcaactatgtatgtagagagctgggtgcgcaagagccttaacagat 4051
 Db 1041 tcatgattgcatcaactatgtatgtagagagctgggtgcgcaagagccttaacagat 1100
 QY 4052 accgtgactgagagagtgcttaacccgttctactgacagagacagctatgttgaaag 4111
 Db 1101 accgtgactgagagagtgcttaacccgttctactgacagagacagctatgttgaaag 1160
 QY 4112 taagagatagctgtggggcgagcagtgtagcagagccgacagatctggcgacagcct 4171
 Db 1161 taagagatagctgtggggcgagcagtgtagcagagccgacagatctggcgacagcct 1220
 QY 4172 gccggaaggaactcaacgcaaaagcgtgactatgagcctcaacagagcttttgagct 4231
 Db 1221 gccggaaggaactcaacgcaaaagcgtgactatgagcctcaacagagcttttgagct 1280
 QY 4232 ggcggaaggaactcaacgcaaaagcgtgactatgagcctcaacagagcttttgagct 4291
 Db 1281 ggcggaaggaactcaacgcaaaagcgtgactatgagcctcaacagagcttttgagct 1340
 QY 4292 cgtacacggggacatctataacaaaggggttgcttaactcagcagggaggaataaa 4351
 Db 1341 cgtacacggggacatctataacaaaggggttgcttaactcagcagggaggaataaa 1400
 QY 4352 gaacaaaggaataatctcaagcctattagaagccttaattgcaaaagagctagctat 4411
 Db 1401 gaacaaaggaataatctcaagcctattagaagccttaattgcaaaagagctagctat 1460
 QY 4412 tatatactgtctgtgacatcagaaagcagaagatctcaatctagaaggaacagatggc 4471
 Db 1461 tatatactgtctgtgacatcagaaagcagaagatctcaatctagaaggaacagatggc 1520

XX W09853104-A2.
 XX 26-NOV-1998.
 XX 18-MAY-1998; 98WO-GB01428.
 XX 16-MAY-1997; 97GB-0010154.
 XX (MED-) MEDICAL RES COUNCIL.
 XX Stoye JP, Weiss RA;
 DR WPI: 1999-045324/04.
 DR P-PSDB: AAW85452.
 XX
 XX Newly isolated nucleic acid probe capable of hybridizing to either
 PT the PERV-A or PERV-B env gene - useful in the detection of
 PT retroviruses, and their subtypes, in a sample of porcine/human
 PT tissue
 XX
 XX Claim 3: Page 21-23; 36pp; English.
 XX
 XX The present sequence encodes a pig endogenous retrovirus (PERV)-B
 CC envelope protein. PERV exists in two different subtypes, PERV-A and
 CC PERV-B. The differences are reflected in sequence divergence in the
 CC (env) genes of PERV-A and PERV-B. The probes and primers are used in
 CC a method to detect retroviruses in a sample of porcine/human tissue,
 CC particularly primary porcine tissue and human cell lines that have been
 CC cultivated in the presence of a porcine cell line, or human tissue from
 CC one of the PERV env genes can also be determined.
 CC
 XX Sequence 3482 BP; 927 A; 854 C; 867 G; 834 T; 0 other;

Query Match 17.9%; Score 1466; DB 20; Length 3482;
 Best Local Similarity 99.38; Pred. No. 0;
 Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

QY 5639 gagcgaacccctccgactcgggtgagagccgaagaagctgagaatcccttaagcttcg 5698
 DB 930 ggcgcacccctccgactcgggtgagagccgaagaagctgagaatcccttaagcttcg 5698
 QY 5699 cctccatcgctggttcttacttacttaacaatactcccaagcagtagtaaacgctta 5758
 DB 990 cctccatcgctggttcttacttacttaacaatactcccaagcagtagtaaacgctta 5758
 QY 5759 taagaacgtcgaacccccaatagaccccttaacccctgctgctgattatgaacgtgata 5818
 DB 1050 taagaacgtcgaacccccaatagaccccttaacccctgctgctgattatgaacgtgata 5818
 QY 5819 cgggtgtcaactgttaataagcaactcgaagtgtgtctctcgaagcaccgtggtgcttgac 5878
 DB 1110 cgggtgtcaactgttaataagcaactcgaagtgtgtctctcgaagcaccgtggtgcttgac 5878
 QY 5879 tgcattctgcgcctcgattgataaaccgcgtgttaaaagcaacactcccaactagcttc 5938
 DB 1170 tgcattctgcgcctcgattgataaaccgcgtgttaaaagcaacactcccaactagcttc 5938
 QY 5939 gtagtatagtgttctatctgctccagcagacagagaagaataactgttggtgcttcg 5998
 DB 1230 gtagtatagtgttctatctgctccagcagacagagaagaataactgttggtgcttcg 5998
 QY 5999 gggaaactctctgttagagatgagctgctgctacactccaaagagagagactggaatggc 6058
 DB 1290 gggaaactctctgttagagatgagctgctgctacactccaaagagagagactggaatggc 6058
 QY 6059 cgatctctccagagaccgggttaaattctcttcttgcactccgcgcggcggaagtaaca 6118
 DB 1350 cgatctctccagagaccgggttaaattctcttcttgcactccgcgcggcggaagtaaca 6118

QY 6119 aaatgataaactatataaagaagagctgctcccaactaagacttagatctataaaga 6178
 DB 1410 aagtgataaactatataaagaagagctgctcccaactaagacttagatctataaaga 6178
 QY 6179 taagtctacg--aaaggaacaggaataatcacaagtgataaagtatgagct 6236
 DB 1470 taagtctacg--aaaggaacaggaataatcacaagtgataaagtatgagct 6236
 QY 6237 ggggaatagttttat-tatagcgggggagcaggggtcccaacttaaccatcgctta 6295
 DB 1530 ggggaatagtttttataataatagcgggggagcaggggtcccaacttaaccatcgctta 6295
 QY 6296 ggaataagagcggggaacagaaacccctgtgcaatggagccgataaagtaactgctgagc 6355
 DB 1590 ggaataagagcggggaacagaaacccctgtgcaatggagccgataaagtaactgctgagc 6355
 QY 6356 aaggggcccccggcggcgggagccacacatacttcgggtgcccgaatgaactcgtcgc 6415
 DB 1650 aaggggcccccggcggcgggagccacacatacttcgggtgcccgaatgaactcgtcgc 6415
 QY 6416 ggcctgacataacacagccgcttagcaacagatgacactggaattatcttcaacacgc 6475
 DB 1710 ggcctgacataacacagccgcttagcaacagatgacactggaattatcttcaacacgc 6475
 QY 6476 ctgaacactcccaaggttctctgttaagaacagagacactctcaagctcaccagc 6535
 DB 1770 ctgaacactcccaaggttctctgttaagaacagagacactctcaagctcaccagc 6535
 QY 6536 gagcttccaagcactcaccacccgacccctgtgcaactctctgttgcctgtgc 6595
 DB 1830 gagcttccaagcactcaccacccgacccctgtgcaactctctgttgcctgtgc 6595
 QY 6596 tatctcagggcctccttatatagagggatgagcttaagaagaataatcaagtgacca 6655
 DB 1890 tatctcagggcctccttatatagagggatgagcttaagaagaataatcaagtgacca 6655
 QY 6656 aagagcagaataacatgatacgtggtggtccgaataagacttaccactcgaagttt 6715
 DB 1950 aagagcagaataacatgatacgtggtggtccgaataagacttaccactcgaagttt 6715
 QY 6716 ccgggaaggggagcagatgataaggaagctcccaactcccaacacacttgcataagta 6775
 DB 2010 ccgggaaggggagcagatgataaggaagctcccaactcccaacacacttgcataagta 6775
 QY 6776 ctgtgtttatgagcagcctccgaataatcagatattatgactgttataacagtgct 6835
 DB 2070 ctgtgtttatgagcagcctccgaataatcagatattatgactgttataacagtgct 6835
 QY 6836 gggcagtgcaatcagtggttaacccctgtgttccacactcagcttcaacacacccaag 6895
 DB 2130 gggcagtgcaatcagtggttaacccctgtgttccacactcagcttcaacacacccaag 6895
 QY 6896 attcgtgtgcaatggtgcgaatcgttcccccgggtgtactacactcgaagaaagtgctcc 6955
 DB 2190 attcgtgtgcaatggtgcgaatcgttcccccgggtgtactacactcgaagaaagtgctcc 6955
 QY 6956 ttgataataatgatactcgtataacacgaacaaagaagaacccgtatccctcctag 7015
 DB 2250 ttgataataatgatactcgtataacacgaacaaagaagaacccgtatccctcctag 7015
 QY 7016 ctgtatgtcgtgattagggagcgcgttggcgttagaagcgggagcgtgctcgtatga 7075
 DB 2310 ctgtatgtcgtgattagggagcgcgttggcgttagaagcgggagcgtgctcgtatga 7075
 QY 7076 caggaccccaagcagccttaagaagaagacttgggtgctcactcgcgcataagagagatc 7135
 DB 2370 caggaccccaagcagccttaagaagaagacttgggtgctcactcgcgcataagagagatc 7135
 QY 7136 tccagccttaagaagacttgggtgcaacttagaagaagctccgacttcttctggaag 7195
 DB 2430 tccagccttaagaagacttgggtgcaacttagaagaagctccgacttcttctggaag 7195
 QY 7196 tggcttcaagaacccgagggagtgatgactgctgttcttaagaagaagtggttatgct 7255

Db 2490 tggctctacagacccggaaggatagatcgtgttctcagaagaagtgtgtatgtg 2549
 QY 7256 cagccttaaaagaatgttctctatgtatgatactcaaggagccatcagagactcca 7315
 Db 2550 cagccttaaaagaatgttctctatgtatgatactcaaggagccatcagagactcca 2609
 QY 7316 tgaacaagcttagaaaaagttagagagcgtcgaagaaggaaagagcttagcagaggt 7375
 Db 2610 ttagcaagcttagaaaaagttagagagcgtcgaagaaggaaagagcttagcagaggt 2669
 QY 7376 ggtttgaagatgttcaacaggtctcctgtgatacgaacccctgtcttctgtcgaag 7435
 Db 2670 ggtttgaagatgttcaacaggtctcctgtgatacgaacccctgtcttctgtcgaag 2729
 QY 7436 ggcctctagatcgtctcgttacttaactagttggtggtctgttcttaataatgtttg 7495
 Db 2730 gacctctagatcgtctcgttacttaactagttggtggtctgttcttaataatgtttg 2789
 QY 7496 ttgctctgttagaagaacgagtgagtgagtcagatcagatgagtaaggaacaagtaac 7555
 Db 2790 ttgctctgttagaagaacgagtgagtgagtcagatcagatgagtaaggaacaagtaac 2849
 QY 7556 aagcctcttagcgaagaagaacttagacttagccttccagcttcttaagataaact 7615
 Db 2850 aagcctcttagcgaagaagaacttagacttagccttccagcttcttaagataaact 2909
 QY 7616 attacaagaagaagaatgtggaatgaaagtgaataatgaacttaaccctccagac 7675
 Db 2910 attacaagaagaagaatgtggaatgaaagtgaataatgaacttaaccctccagac 2969
 QY 7676 cgaagaagttaataaagccttaaatgtcccccaattacagaccctgtgtgtcagat 7735
 Db 2970 cgaagaagttaataaagccttaaatgtcccccaattacagaccctgtgtgtcagat 3029
 QY 7736 aatatgtagaagtgatcacacttctctatgttccagaggtctcttccgtgactaagtaag 7795
 Db 3030 aatatgtagaagtgatcacacttctctatgttccagaggtctcttccgtgactaagtaag 3089
 QY 7796 ataaagaagaatgagttgactaatcgtctatctgtlaaactgtactgtgacacat 7855
 Db 3090 ataaagaagaatgagttgactaatcgtctatctgtlaaactgtactgtgacacat 3149
 QY 7856 agaagaattgatacatattgacagcctagtgacttatcacaatgtaactcttcaactc 7915
 Db 3150 agaagaattgatacatattgacagcctagtgacttatcacaatgtaactcttcaactc 3209
 QY 7916 tggccagagggccacagcagatgagacccctcagagctatatttaaatgattgtccacg 7975
 Db 3210 tggccagagggccacagcagatgagacccctcagagctatatttaaatgattgtccacg 3269
 QY 7976 agcgcgggtctcgtatattttaaatgattgttccatgtagcgcggtc-tcgaattt 8034
 Db 3270 agcgcgggtctcgtatattttaaatgattgttccatgtagcgcggtc-tcgaattt 8094
 QY 8035 taaatgattgttgtgagcagcaggttgtgtgtaaccccatataaagcgtgtccgat 8094
 Db 3330 taaatgattgttgtgagcagcaggttgtgtgtaaccccatataaagcgtgtccgat 3389
 QY 8095 tccgcactcgggcgagcagctctacacccctgcgtgtgtgtaagcagctgtggtcccgccg 8154
 Db 3390 tccgcactcgggcgagcagctctacacccctgcgtgtgtgtaagcagctgtggtcccgccg 3449
 QY 8155 gcttggataaaatcctcttctgtgtgtgcatc 8187
 Db 3450 gcttggataaaatcctctctgcgttgcac 3482

RESULT 8
 AAC67020
 ID AAC67020 standard: DNA: 4402 BP.
 XX
 AC AAC67020;

XX 27-MAR-2001 (first entry)
 DT
 XX PERV env protein coding sequence SEQ ID NO: 20.
 DE
 XX Xenotransplantation; infectious agent; vaccine; ds.
 XX
 XX Porcine endogenous retrovirus.
 OS
 XX MO200071726-A1.
 PN
 XX 30-NOV-2000.
 PD
 XX 24-MAY-2000; 2000MO-US14296.
 PF
 XX 24-MAY-1999; 99US-0135631.
 PR
 XX (MAYO-) MAYO MEDICAL VENTURES.
 PA
 XX Federpsiel MJ;
 PI
 XX WPI: 2001-032041/04.
 DR
 XX
 XX Inhibiting or preventing infectious agent transmission in mammalian
 PT transplant recipients, by introducing recombinant DNA comprising DNA
 PT encoding extracellular proteins of the agent into donor cells, such as
 CC swine cells -
 CC
 CC Claim 16; Page 105-106; 144pp: English.
 PS
 XX
 XX The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 CC
 CC Sequence 4402 BP; 1259 A; 1085 C; 1111 G; 947 T; 0 other;
 SQ
 Query Match 10.3%; Score 848; DB 22; Length 4402;
 Best Local Similarity 98.8%; Pred. NO. 0;
 Matches 2198; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 QY 2982 ttcctcgtgaattacacccctagcccaacatttgccttcgaatgagagatcca 3041
 Db 1043 ttctgcttagatatacaccctcagccacacatttgccttcgaatgagagatcca 1102
 QY 3042 ggtacgggaagaacccgagcagctacacttgagccgactgtcccaagggttcaagaactcc 3101
 Db 1103 ggtacgggaagaacccgagcagctacacttgagccgactgtcccaagggttcaagaactcc 1162
 QY 3102 ccgaacacatttgaagaaagccctacacagagacccctggccaacttcaagatccaacacct 3161
 Db 1163 ccgaacacatttgaagaaagccctacacagagacccctggccaacttcaagatccaacacct 1222
 QY 3162 caagtgacccctctccagtagtcgtagtgacactgtcttgcgggagccacaacagagac 3221
 Db 1223 caagtgacccctctccagtagtcgtagtgacactgtcttgcgggagccacaacagagac 1282
 QY 3222 tgcctagaagtagcaagaagcactacgtctggaatttctgacctaggtctacagacctct 3281
 Db 1283 tgcctagaagtagcaagaagcactacgtctggaatttctgacctaggtctacagacctct 1342
 QY 3282 gctagaagaagccagacttgcagagagagaggttaactacttgggttacaagtttcggggc 3341
 Db 1343 gctagaagaagccagacttgcagagagagaggttaactacttgggttacaagtttcggggc 1402
 QY 3342 gggcagcgaatggtctgacggaagcagcgaagaagaactgtagtcaagataccggcccaacc 3401
 Db 1403 gggcagcgaatggtctgacggaagcagcgaagaagaactgtagtcaagataccggcccaacc 1462
 QY 3402 acagcacaagaatgagagagtttcttgggagcagctggaatttgcagactgtgagatccg 3461

Db 1463 aacgccaacaaagtgaagagatttttggggaacgtgtgatttgcagactgtgatcccg 1522
 QY 3462 ggggttgcgaccttagcagcccaactacccggttaacccaagaagaagggtatcttc 3521
 Db 1523 ggggttgcgaccttagcagcccaactacccggttaacccaagaagaagggtatcttc 3521
 QY 3522 tgggtctctggccacccaagaaggcatcttgatgctacaaaaggccctgcgacgcacac 3581
 Db 1583 tgggtctctggccacccaagaaggcatcttgatgctacaaaaggccctgcgacgcacac 3581
 QY 3582 gctctggccctccctgcgacgttaacccaactaccccttaaccccttaaccccttaaccc 3641
 Db 1643 gctctggccctccctgcgacgttaacccaactaccccttaaccccttaaccccttaaccc 3641
 QY 3642 gtagcccgagaggttttaacccaaccccttagagacccatgagagagacgtgtgcacac 3701
 Db 1703 gtagcccgagaggttttaacccaaccccttagagacccatgagagagacgtgtgcacac 3701
 QY 3702 tcaagaagacttgatcccttagagacccatgagagacccatgagagacccatgagagac 3761
 Db 1763 tcaagaagacttgatcccttagagacccatgagagacccatgagagacccatgagagac 3761
 QY 3762 gtggccacttgatcgaagacgcgtgaacaaattgactttggagacagataataactgata 3821
 Db 1823 gtggccacttgatcgaagacgcgtgaacaaattgactttggagacagataataactgata 3821
 QY 3822 gcccccactgcatggagagacatggttgcgaccccccagacccagacccagacccagaccc 3881
 Db 1883 gcccccactgcatggagagacatggttgcgaccccccagacccagacccagacccagaccc 3881
 QY 3882 cgcataacccactataacagccctgtcttcacataagagaggtacacttgcgcacacaccc 3941
 Db 1943 cgcataacccactataacagccctgtcttcacataagagaggtacacttgcgcacacaccc 3941
 QY 3942 gctcctaaccgacacacttcttgcgcacacacacacacacacacacacacacacacacac 4001
 Db 4003 gctcctaaccgacacacttcttgcgcacacacacacacacacacacacacacacacacac 4001
 QY 4002 catcaactatctgataagagacgtgggttcgcgacagagaccccttaacacacacacacac 4061
 Db 2063 catcaactatctgataagagacgtgggttcgcgacagagaccccttaacacacacacacac 4061
 QY 4062 ggaagaaggtgaacactgttactgaacggaagacacacacacacacacacacacacacac 4121
 Db 2123 ggaagaaggtgaacactgttactgaacggaagacacacacacacacacacacacacacac 4121
 QY 4122 gctggggcgaggtggtgagacgggacccgacacacacacacacacacacacacacacacac 4181
 Db 2183 gctggggcgaggtggtgagacgggacccgacacacacacacacacacacacacacacacac 4181
 QY 4182 acttcagcgcaaaagcgtgacgtcactgaacggaacacacacacacacacacacacacacac 4241
 Db 2243 acttcagcgcaaaagcgtgacgtcactgaacggaacacacacacacacacacacacacacac 4241
 QY 4242 aaatccataaacttatagac 4301
 Db 2303 aaatccataaacttatagac 4301
 QY 4302 gccatctataaacaagaagggtgtgcttactcagcagagagaggaataaagaacaagaag 4361
 Db 2363 gccatctataaacaagaagggtgtgcttactcagcagagagaggaataaagaacaagaag 4361
 QY 4362 gaaatcttaagccttatagaagccttaacattgccaanaaaggctagactatataactgt 4421
 Db 2423 gaaatcttaagccttatagaagccttaacattgccaanaaaggctagactatataactgt 4421
 QY 4422 cctggaacatcagaagaacccaagaatcctacatctagaaggaacacagatgctgcagcgggt 4481
 Db 2483 cctggaacatcagaagaacccaagaatcctacatctagaaggaacacagatgctgcagcgggt 4481
 QY 4482 gccaaagcagcagccagcgtgttaacacttgccttaataaagaagcccaagcccca 4541
 Db 2543 gccaaagcagcagccagcgtgttaacacttgccttaataaagaagcccaagcccca 2602

QY 4542 gaaccacagacagctacaccccttagaagactggcagaagagataaaaaagataagaccagctc 4601
 Db 2603 gaaccacagacagctacaccccttagaagactggcagaagagataaaaaagataagaccagctc 4601
 QY 4602 tctgaagactccggaggggagacctgtctacacacacacacacacacacacacacacacacac 4661
 Db 2663 tctgaagactccggaggggagacctgtctacacacacacacacacacacacacacacacacac 4661
 QY 4662 gaaaggttaagatatgtccac 4721
 Db 2723 gaaaggttaagatatgtccac 4721
 QY 4722 cagcagcttgatcagaac 4781
 Db 2783 cagcagcttgatcagaac 4781
 QY 4782 gtgttcaaacatgtgttgcaccttgccacacacacacacacacacacacacacacacacacac 4841
 Db 2843 gtgttcaaacatgtgttgcaccttgccacacacacacacacacacacacacacacacacacac 4841
 QY 4842 ggaagaagactaagggggaagccac 4901
 Db 2903 ggaagaagactaagggggaagccacacacacacacacacacacacacacacacacacacac 4901
 QY 4902 aagccggttaataacaggaac 4961
 Db 2963 aagccggttaataacaggaac 4961
 QY 4962 gtagagacttaccac 5021
 Db 3023 gtagagacttaccac 5021
 QY 5022 gaaattttccaaagattgggaatacctaaagglaaagggttaagacacacacacacacacac 5081
 Db 3083 gaaattttccaaagattgggaatacctaaagglaaagggttaagacacacacacacacacac 5081
 QY 5082 gttgccaggtgaagtcagacagac 5141
 Db 3143 gttgccaggtgaagtcagacagac 5141
 QY 5142 gcatcagaaccccaagcgtcagacagacacacacacacacacacacacacacacacacacac 5201
 Db 3203 gcatcagaaccccaagcgtcagacagacacacacacacacacacacacacacacacacacac 5201
 QY 5202 cttaac 5206
 Db 3263 cttaac 3267

RESULT 9
 AAC67021
 ID AAC67021 standard; DNM: 6076 BP.
 AC AAC67021:
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE PERV env protein coding sequence SEQ ID NO: 21.
 XX
 KM Xenotransplantation; infectious agent; vaccine; ds.
 XX
 OS Porcine endogenous retrovirus.
 PN WO200071726-A1.
 PD 30-NOV-2000.
 XX
 PF 24-MAY-2000; 2000WO-US14296.
 XX
 PR 24-MAY-1999; 99US-0135631.
 XX
 PA (MAYO-) MAYO MEDICAL VENTURES.

XX Federpriel MJ:
 XX WPI: 2001-032041/04.
 XX Inhibiting or preventing infectious agent transmission in mammalian
 XX transplant recipients, by introducing recombinant DNA comprising DNA
 XX encoding extracellular proteins of the agent into donor cells, such as
 XX swine cells -
 XX
 XX Claim 16: Page 107-109; 144pp; English.
 XX
 XX The present invention provides a method to prevent the transmission of
 XX infectious agents during xenotransplantation. This involves introducing
 XX to donor swine cells a recombinant DNA encoding a peptide fragment from
 XX the infectious agent, and then introducing these cells into the
 XX transplant recipient.
 XX
 XX Sequence 6076 BP: 1613 A; 1512 C; 1609 G; 1342 T; 0 other:

Query Match 9.7%; Score 797; DB 22; Length 6076;
 Best Local Similarity 98.7%; Pred. No. 1.8e-297;
 Matches 2197; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 2982 tctgtcctgagattacacccactagccacacttctgtccttcgaatgagagatcca 3041
 Db 3577 tctgtcctgagattacacccactagccacacttctgtccttcgaatgagagatcca 3636
 QY 3042 ggtacgaggaagacggcgagctcaactcgtgaccccgcccaagggttaagaactcc 3101
 Db 3637 ggtacgaggaagacggcgagctcaactcgtgaccccgcccaagggttaagaactcc 3696
 QY 3102 ccgacacactttagcaagacccctacacagagacctgagccaaacttaagaaatcaaacctc 3161
 Db 3697 ccgacacactttagcaagacccctacacagagacctgagccaaacttaagaaatcaaacctc 3756
 QY 3162 caagtacacctcctccagtaactgagatcgtcctcgtcgcgagagccacacaaagagac 3221
 Db 3757 caagtacacctcctccagtaactgagatcgtcctcgtcgcgagagccacacaaagagac 3816
 QY 3222 tggcttgaagatgacgaagacactacgtcgtgaattctcgaactagctacagagctctc 3281
 Db 3817 tggcttgaagatgacgaagacactacgtcgtgaattctcgaactagctacagagctctc 3876
 QY 3282 gctaaagaagcccaagatttgcagagagagatataacttgggttacagtttgcggggc 3341
 Db 3877 gctaaagaagcccaagatttgcagagagagatataacttgggttacagtttgcggggc 3936
 QY 3342 gggcagagatgagctgagagagagacggaagaagaactgtagtccagaataccgacccaaac 3401
 Db 3937 gggcagagatgagctgagagagagacggaagaagaactgtagtccagaataccgacccaaac 3996
 QY 3402 aacacccaacaagatgagagagatttgggagacgtggaatttgcagactgtgagaccg 3461
 Db 3997 aacacccaacaagatgagagagatttgggagacgtggaatttgcagactgtgagaccg 4056
 QY 3462 ggggttgcaccttaagacgcccacttaccgctaacccaagaagaagggtgattctcc 3521
 Db 4057 ggggttgcaccttaagacgcccacttaccgctaacccaagaagaagggtgattctcc 4116
 QY 3522 tgggtctctgagacccagaaagcatttgatgatacaaaagagccctgtgagacacact 3581
 Db 4117 tgggtctctgagacccagaaagcatttgatgatacaaaagagccctgtgagacacact 4176
 QY 3582 gctctgacctcctgacgtataactaaaccttacccttaattggtgagtgagagga 3641
 Db 4177 gctctgacctcctgacgtataactaaaccttacccttaattggtgagtgagagga 4236
 QY 3642 gtagcccgagagagtttaccacaaaccttagacatagagagagacgtgttgcctactcg 3701
 Db 4237 gtagcccgagagagtttaccacaaaccttagacatagagagagacgtgttgcctactcg 4296

QY 3702 tcaagaagcttgatccttgaagcagtggtgcccgtatgtctgaagctacgcagct 3761
 Db 4297 tcaagaagcttgatccttgaagcagtggtgcccgtatgtctgaagctacgcagct 4356
 QY 3762 gtggcatalactgtcaagagacgtcagcaaatgacttgggagacatatataactgttaata 3821
 Db 4357 gtggcatalactgtcagagacgtcagcaaatgacttgggagacatatataactgttaata 4416
 QY 3822 gcccccagatctggagagacatgtctcgcagaccccccagacagatgagatcaacagcc 3881
 Db 4417 gcccccagatctggagagacatgtctcgcagaccccccagacagatgagatcaacagcc 4476
 QY 3882 cgcatagacccataccaagcctgtctctacagagaggttcaacttgcctccacagcc 3941
 Db 4477 cgcatagacccataccaagcctgtctctacagagaggttcaacttgcctccacagcc 4536
 QY 3942 gctctcaacctgcacactctcgtcgtgaagagagcttgaacagatgatactgttc 4001
 Db 4537 gctctcaacctgcacactctcgtcgtgaagagagcttgaacagatgatactgttc 4596
 QY 4002 catcaactatgattgaagagacgtgggtccgcagagaccttaacagacataccgtact 4061
 Db 4597 catcaactatgattgaagagacgtgggtccgcagagaccttaacagacataccgtact 4656
 QY 4062 ggaagagtgctaacctgttcaactgacggaagcagctatgtgtgagaggttaagagatg 4121
 Db 4657 ggaagagtgctaacctgttcaactgacggaagcagctatgtgtgagaggttaagagatg 4716
 QY 4122 gctggggcgagctgtgtgagacggacccgcagacactggtggcacaagacctgcggagag 4181
 Db 4717 gctggggcgagctgtgtgagacggacccgcagacactggtggcacaagacctgcggagag 4776
 QY 4182 acttcagccaaaagccttagctatctatggccctacgcgaagcttgcggctgcggagag 4241
 Db 4777 acttcagccaaaagccttagctatctatggccctacgcgaagcttgcggctgcggagag 4836
 QY 4242 aaatccataaacttatatacgcagacagagatgtccttgcagctgcagctacacgg 4301
 Db 4837 aaatccataaacttatatacgcagacagagatgtccttgcagctgcagctacacgg 4896
 QY 4302 gccatctataaacaaggggtgtgtcttaacctacagcagggaggaataaagaacaagag 4361
 Db 4897 gccatctataaacaaggggtgtgtcttaacctacagcagggaggaataaagaacaagag 4956
 QY 4362 gaaatttcaagcctatagaagccttaacttggccaaaaggctagctattatacgt 4421
 Db 4957 gaaatttcaagcctatagaagccttaacttggccaaaaggctagctattatacgt 5016
 QY 4422 ccttgagacatagaagaagccaagaatcctacatctagaagggagacagatggtgacccggt 4481
 Db 5017 ccttgagacatagaagaagccaagaatcctacatctagaagggagacagatggtgacccggt 5076
 QY 4482 gccaaagcagccagccagctgttaacctctcgtcctataatgaaacgcgccaagaccca 4541
 Db 5077 gccaaagcagccagccagctgttaacctctcgtcctataatgaaacgcgccaagaccca 5136
 QY 4542 gaaacccagacagctagacacccctagaagagcttgcagaagagataaagaagtagaccagttc 4601
 Db 5137 gaaacccagacagctagacacccctagaagagcttgcagaagagataaagaagtagaccagttc 5196
 QY 4602 tctgagactccgggggggagctgtctacactcatabatgggaaggaatcctgtcccccacaa 4661
 Db 5197 tctgagactccgggggggagctgtctacactcatabatgggaaggaatcctgtcccccacaa 5256
 QY 4662 gaaaggttaagaatatgtcccaagaatagatcgttcaacccacttagaagactaaacacgtg 4721
 Db 5257 gaaaggttaagaatatgtcccaagaatagatcgttcaacccacttagaagactaaacacgtg 5316
 QY 4722 cagcaggttggtcagaacalcccttatactatgtcttgaagctaccagagatgtgctgactcg 4781
 Db 5317 cagcaggttggtcagaacalcccttatactatgtcttgaagctaccagagatgtgctgactcg 5376
 QY 4782 gtgttcaaacatltgtgtcctgtcagctgtgttaatgtcttaactctccagaataacctca 4841

Query Match	9.7%	Score 797	DB 22	Length 7362
Best Local Similarity	98.7%	Pred. No. 1.8e-297		
Matches 2197; Conservative	0	Mismatches 28	Indels 0	Gaps 0

[illegible]

Db 4602 gctctaacccgtccactcttctgctgaagaactgtatgaaccagtgactatgaltgc 4661
 QY 4002 catcaactatgtatgagagactgggttcgcaagacacttacagacatccgtact 4061
 Db 4662 catcaactatgtatgagagactgggttcgcaagacacttacagacatccgtact 4721
 QY 4062 ggggaagtgtaactcgtgtcactgaacgaactatgtgtgtaaggttaagagatg 4121
 Db 4722 ggaagagtgtaactcgtgtcactgaacgaactatgtgtgtaaggttaagagatg 4781
 QY 4122 gctggggcgagctgggtggaagggacccgacagatctgggacagacgtcgcggaaaga 4181
 Db 4782 gctggggcgagctgggtggaagggacccgacagatctgggacagacgtcgcggaaaga 4841
 QY 4182 acttcagcgcaaaagctgagctcagtcgcccctacagcaagcttcgcgctgcgaaagg 4241
 Db 4842 acttcagcgcaaaagctgagctcagtcgcccctacagcaagcttcgcgctgcgaaagg 4901
 QY 4242 aaatccataacattatatacgaacagagatgctcttgagactgacagctaacaggg 4301
 Db 4902 aaatccataacattatatacgaacagagatgctcttgagactgacagctaacaggg 4961
 QY 4302 gccatctataaacaaggggtgtgctactctcagcgagggggaaataagaacaagaag 4361
 Db 4962 gccatctataaacaaggggtgtgctactctcagcgagggggaaataagaacaagaag 5021
 QY 4362 gaaattcgaagcctattagaaagccttactgtccaaagagcttagctatatacctgt 4421
 Db 5022 gaaattcgaagcctattagaaagccttactgtccaaagagcttagctatatacctgt 5081
 QY 4422 cctggacatacgaagaagccaaagatctatctagaaggaacagatggcttgacgggt 4481
 Db 5082 cctggacatacgaagaagccaaagatctatctagaaggaacagatggcttgacgggt 5141
 QY 4482 gccaaagcagagccagccagcgttaacctctgctataatagaacgccaaagccca 4541
 Db 5142 gccaaagcagagccagccagcgttaacctctgctataatagaacgccaaagccca 5201
 QY 4542 gaaccacagacagatatacccttagaagaactggcagaagataaaagatagacagttc 4601
 Db 5202 gaaccacagacagatatacccttagaagaactggcagaagataaaagatagacagttc 5261
 QY 4602 tctgaagcctcggaggggagcctgtctatacctatatagggaaggaatcctgccccacaa 4661
 Db 5262 tctgaagcctcggaggggagcctgtctatacctatatagggaaggaatcctgccccacaa 5321
 QY 4662 gaaggggttagaatatgtccacaagatatacctgttcaaccacttagaactcaaacactg 4721
 Db 5322 gaaggggttagaatatgtccacaagatatacctgttcaaccacttagaactcaaacactg 5381
 QY 4722 cagcaggttgtaagaacatcccttactatgttctgagggctacagaggtggtactcg 4781
 Db 5382 cagcaggttgtaagaacatcccttactatgttctgagggctacagaggtggtactcg 5441
 QY 4782 gtgtgaacaaatgtgtgcccctgcagcgtgttaatgttaactcctccagaatattcca 4841
 Db 5442 gtgtgaacaaatgtgtgcccctgcagcgtgttaatgttaactcctccagaatattcca 5501
 QY 4842 ggaagaagactaaagggaagccacccagcgctcactggaaggttagacttaactgagta 4901
 Db 5502 ggaagaagactaaagggaagccacccagcgctcactggaaggttagacttaactgagta 5561
 QY 4902 aaagcggctaaatacggaaacaataatctatgtgttttgtaagaccccttcaagatgg 4961
 Db 5562 aaagcggctaaatacggaaacaataatctatgtgttttgtaagaccccttcaagatgg 5621
 QY 4962 gttagagcttaactactaagaagaagacttcaacggtgtgtgcttaagaataactacggag 5021
 Db 5622 gttagagcttaactactaagaagaagacttcaacggtgtgtgcttaagaataactacggag 5681
 QY 5022 gaaattttccaagattggaatacctaaggttaataggtcagaacatgtgtccagcttc 5081
 Db 5682 gaaattttccaagattggaatacctaaggttaataggtcagaacatgtgtccagcttc 5741

QY 5082 gtgtgccaggttaagtcagaggaactggccaagatatgtgggattgtatggaaactgcatgt 5141
 Db 5742 gtgtgccaggttaagtcagaggaactggccaagatatgtgggattgtatggaaactgcatgt 5801
 QY 5142 gcatagaagcccaagagctcagagacaggttagagagatgatatagaacattaaagagacc 5201
 Db 5802 gcatagaagcccaagagctcagagacaggttagagagatgatatagaacattaaagagacc 5861
 QY 5202 cttaac 5206
 Db 5862 cttaac 5866

RESULT 11

AAC67032 standard; DNA; 6076 BP.

ID AAC67032

AAC67032: (first entry)

27-MAR-2001 (first entry)

PERV env protein coding sequence seq ID NO: 32.

Xenotransplantation; infectious agent; vaccine; ds.

Porcine endogenous retrovirus.

WO200071726-A1.

30-NOV-2000.

24-MAY-2000; 2000WO-US14296.

24-MAY-1999; 99US-0135631.

(MAYO-) MAYO MEDICAL VENTURES.

Federspiel MJ;

WPI; 2001-032041/04.

Inhibiting or preventing infectious agent transmission in mammalian

transplant recipients, by introducing recombinant DNA comprising DNA

encoding extracellular proteins of the agent into donor cells, such as

swine cells

Claim 16; Page 117-119; 144pp; English.

The present invention provides a method to prevent the transmission of

infectious agents during xenotransplantation. This involves introducing

to donor swine cells a recombinant DNA encoding a peptide fragment from

the infectious agent, and then introducing these cells into the

transplant recipient.

Sequence 6076 BP; 1619 A; 1505 C; 1598 G; 1354 T; 0 other;

Query Match 9.1%; Score 746; DB 22; Length 6076;

Best Local Similarity 98.7%; Pred No. 7.7e-278;

Matches 2196; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

2982 ttctgctgagatatacaccacactagcaacacacttttgccttcgaatggagagatcca 3041

Db 3577 ttctgctgagatatacaccacactagcaacacacttttgccttcgaatggagagatcca 3636

QY 3042 ggtacgggaagacccggagcgtctactgtgaccggagctgcccgaaggttcaagaactcc 3101

Db 3637 ggtacgggaagacccggagcgtctactgtgaccggagctgcccgaaggttcaagaactcc 3696

QY 3102 ccgacacatttgaacgaagccctacacagaggaactgtgccaacttcagatccaacacct 3161

Db 3697 ccgacacatttgaacgaagccctacacagaggaactgtgccaacttcagatccaacacct 3756

AC AAT74884;
 XX 09-FEB-1998 (first entry)
 DE Miniature swine retrovirus cDNA.
 XX Retrovirus; porcine; GAG protein; POL protein; ENV protein;
 KM xenotransplantation; infectious; provirus; organ transplant; donor;
 KW activated virus; PCR; ss.
 XX Porcine retrovirus.
 OS
 XX Key Location/Qualifiers
 FH CDS 585..2159
 FT /*tag= a
 FT 585..2156
 FT mat_peptide
 FT /*tag= b
 FT /note= "putative GAG protein"
 FT 2307..5744
 FT CDS
 FT /*tag= c
 FT 2307..5741
 FT mat_peptide
 FT /*tag= d
 FT /note= "putative POL protein"
 FT 5620..7536
 FT CDS
 FT /*tag= e
 FT 5620..7533
 FT mat_peptide
 FT /*tag= f
 FT /note= "putative ENV protein"
 XX
 XX MO9721836-A1.
 XX 19-JUN-1997.
 XX 13-DEC-1996; 96MO-US19680.
 XX 14-DEC-1995; 95US-0572645.
 XX (GENO) GEN HOSPITAL CORP.
 XX Fishman JA;
 XX WPI: 1997-332804/30.
 DR P-PSDB: AAM32096-W32098.
 XX
 XX New nucleic acid from porcine retroviruses - used for detecting
 PT viruses in transplant or other tissue and for assessing risk of
 PT transmitting infection to graft recipient
 XX
 PS Claim 22; Fig 3; 128bp; English.
 XX
 XX This cDNA sequence represents a porcine retrovirus from miniature swine
 CC containing the putative coding regions for viral GAG, POL, and ENV
 CC proteins. This sequence and PCR fragments generated from it
 CC (see AAT74812-T74882) can be used to screen organs for the presence of
 CC porcine retroviruses prior to xenotransplantation. Transplantation can
 CC increase the likelihood of retroviral activation if intact and
 CC infectious proviruses are present. The porcine retroviral sequence can be
 CC used to generate probes to determine the level (e.g. copy number) of
 CC intact (i.e. potentially replicating) porcine provirus sequences in a
 CC strain of xenograft transplantation donors. It can be used to detect
 CC mutations, genetic lesions or viral recombinants and also to determine
 CC the histological localization of activated retroviruses. Using Polymerase
 CC Chain Reaction DNA Quantitation (PQD) on blood mononuclear cells,
 CC infectivity titration and susceptibility testing can be performed.
 CC Ultimately animal donors without intact porcine retroviral sequences or a
 CC lower copy number of viral elements could be selected.
 XX
 XX Sequence 7892 BP; 2191 A; 1915 C; 1980 G; 1806 T; 0 other;

Query Match 8.6%; Score 704; DB 18; Length 7892;
 Best Local Similarity 98.7%; Pred. No. 1,1e-261;
 Matches 2204; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY	2973	gattgcttcttctgcttagattacaccccaactagccaaccacttctgcttgcgaatg	3032
DB	2970	gattgcttcttctgcttagattacaccccaactagccaaccacttctgcttgcgaatg	3029
QY	3033	agagatccagtgtaggaagaacccggcagctacactgagccgaactgccccaaaggttc	3092
DB	3030	agagatccagtgtaggaagaacccggcagctacactgagccgaactgccccaaaggttc	3089
QY	3093	aagaactcccgacacatctttagcagaagccctaaccaagaagccctggcccaactcagatc	3152
DB	3090	aagaactcccgacacatctttagcagaagccctaaccaagaagccctggcccaactcagatc	3149
QY	3153	caaacacctcaagtgtagccctccctccagtagcagtagcagctgctgaggaggagccac	3212
DB	3150	caaacacctcaagtgtagccctccctccagtagcagtagcagctgctgaggaggagccac	3209
QY	3213	aaacaggaactgcttaagaagtagcagaagcactacactgctggaattctgactaagctac	3272
DB	3210	aaacaggaactgcttaagaagtagcagaagcactacactgctggaattctgactaagctac	3269
QY	3273	agagccctgcttaagaagaagccagatttgcagaagaagtagtaacttgggttagact	3332
DB	3270	agagccctgcttaagaagaagccagatttgcagaagaagtagtaacttgggttagact	3329
QY	3333	ttgcgggagcgagcagatgctgacgagagcagcaggaagaaactgtagtccagataccg	3392
DB	3330	ttgcgggagcgagcagatgctgacgagagcagcaggaagaaactgtagtccagataccg	3389
QY	3393	gccccaacccaagccaagaagtaggaagatttggggagcagctgagatttgcagactg	3452
DB	3390	gccccaacccaagccaagaagtaggaagatttggggagcagctgagatttgcagactg	3449
QY	3453	tgaatcccgaggttgcagacttagcagcccaactacccgttaacccaagaagaaggg	3512
DB	3450	tgaatcccgaggttgcagacttagcagcccaactacccgttaacccaagaagaaggg	3509
QY	3513	ggattctctggtgctcctgagcaccagaagatcttgatgctatacaaaaagccctgctg	3572
DB	3510	ggattctctggtgctcctgagcaccagaagatcttgatgctatacaaaaagccctgctg	3569
QY	3573	agcgcaactgcttctggtccctccctgacgttaacttaaaccttacccttattgtgtagag	3632
DB	3570	agcgcaactgcttctggtccctccctgacgttaacttaaaccttacccttattgtgtagag	3629
QY	3633	cgtaaggaagtagcccgagaggttttaacccaacccctagagaccatgagagagactgtt	3692
DB	3630	cgtaaggaagtagcccgagaggttttaacccaacccctagagaccatgagagagactgtt	3689
QY	3693	gctactcgttaagaagcttgatccctgttagccagtggttggcccgatgtctgaagct	3752
DB	3690	gctactcgttaagaagcttgatccctgttagccagtggttggcccgatgtctgaagct	3749
QY	3753	atcgagctgtggtcagatctgtaacgaagccttgacaatattgacttgggacagaatata	3812
DB	3750	atcgagctgtggtcagatctgtaacgaagccttgacaatattgacttgggacagaatata	3809
QY	3813	actgtaataagcccccatgacttggagagacatggttggagagcccccagacagatgagt	3872
DB	3810	actgtaataagcccccatgacttggagagacatggttggagagcccccagacagatgagt	3869
QY	3873	accgaagcccgatataccaactatcaaaagctgtcttcaagaagaggttaacttgcgt	3932
DB	3870	accgaagcccgatataccaactatcaaaagctgtcttcaagaagaggttaacttgcgt	3929
QY	3933	ccacgaagccgtctcaaaccttgcacttctgctgtagaagagatgttgtaacagtagact	3992
DB	3930	ccacgaagccgtctcaaaccttgcacttctgctgtagaagagatgttgtaacagtagact	3989
QY	3993	catgattgcatcaacttgaattagaagagcttgggtgcgcaagagacttacagacata	4052
DB	3990	catgattgcatcaacttgaattagaagagcttgggtgcgcaagagacttacagacata	4049

Db 3090 aagaactcccgacacattcttgaagagccctacacagagccctgagccaaacttcagatc 3149
QY 3153 caaacacctcaagttgaacctctccacagtaagttgaagacctgtctcgccggagagcacc 3212
Db 3150 caaacacctcaagttgaacctctccacagtaagttgaagacctgtctcgccggagagcacc 3209
QY 3213 aaacagagctgttagaaggtacgaagagcactactgctggagatgtctgacactagctac 3272
Db 3210 aaacagagctgttagaaggtacgaagagcactactgctggagatgtctgacactagctac 3269
QY 3273 agaacctctgttagaagagccagatcttcagagagagaggttaacatacttggggtaacagt 3332
Db 3270 agaacctctgttagaagagccagatcttcagagagagaggttaacatacttggggtaacagt 3329
QY 3333 ttgcggggcgagcagcagtgctgacgagagcagcagaaagaaacgttgatccagataccg 3392
Db 3330 ttgcggggcgagcagcagtgctgacgagagcagcagaaagaaacgttgatccagataccg 3389
QY 3393 gccccaacacagcccaacaaagttgaagagatcttcggagagcagcttgatcttcagactg 3452
Db 3390 gccccaacacagcccaacaaagttgaagagatcttcggagagcagcttgatcttcagactg 3449
QY 3453 tggatcccggggtttgacagacttagcagcccaactacccgttaacccaagaaagagag 3512
Db 3450 tggatcccggggtttgacagacttagcagcccaactacccgttaacccaagaaagagag 3509
QY 3513 ggaattccctggtccctgagcagcagcagagcagatcttgaatacgaagagccctgtg 3572
Db 3510 ggaattccctggtccctgagcagcagcagagcagatcttgaatacgaagagccctgtg 3569
QY 3573 agagcagctgtctgtgacctccctgaagcgttaacaaaccttaccctttagtggatag 3632
Db 3570 agagcagctgtctgtgacctccctgaagcgttaacaaaccttaccctttagtggatag 3629
QY 3633 cgttaagagagtagcccgagaggttttaacccaacccatgaagagcagatctgtc 3692
Db 3630 cgttaagagagtagcccgagaggttttaacccaacccatgaagagcagatctgtc 3689
QY 3693 gcttacctgtcgaagagagcttgaatccctgaagcagtggtgtgacctgtatgtgaagagct 3752
Db 3690 gcttacctgtcgaagagagcttgaatccctgaagcagtggtgtgacctgtatgtgaagagct 3749
QY 3753 atcgagagctgtgacatactgttgaagagcgttgaacaaatgaacttggagcagaatata 3812
Db 3750 atcgagagctgtgacatactgttgaagagcgttgaacaaatgaacttggagcagaatata 3809
QY 3813 actttaagcccccaatgaatgtgaagacatcgttgcgagcccccgagccgagatgagtg 3872
Db 3810 actttaagcccccaatgaatgtgaagacatcgttgcgagcccccgagccgagatgagtg 3869
QY 3873 accaagcccgatgacccacatacaagcctgtgcttcaacagagaggttcaacttgcct 3932
Db 3870 accaagcccgatgacccacatacaagcctgtgcttcaacagagaggttcaacttgcct 3929
QY 3933 ccacagagcgtcttcaacccctgtcacctctgtcctgaagagagcttgaacaaatgagct 3992
Db 3930 ccacagagcgtcttcaacccctgtcacctctgtcctgaagagagcttgaacaaatgagct 3989
QY 3993 catgattgcatacaactatgtattgaagagagctggggtccgcaagagaccttaccagacata 4052
Db 3990 catgattgcatacaactatgtattgaagagagctggggtccgcaagagaccttaccagacata 4049
QY 4053 ccgctgactgtgaagagtgataactgttcaactgaacggaagagcagctatgtgtgaagtg 4112
Db 4050 ccgctgactgtgaagagtgataactgttcaactgaacggaagagcagctatgtgtgaagtg 4109
QY 4113 aagagagatgctgtggcgagcagtggtgaacgagcccgagatctggcgacagcagctg 4172
Db 4110 aagagagatgctgtggcgagcagtggtgaacgagcccgagatctggcgacagcagctg 4169
QY 4173 ccggaagagacttcagcgaagagctgaagctcagtgccctcagcagagcttggcgctg 4232
Db 4170 ccggaagagacttcagcgaagagctgaagctcagtgccctcagcagagcttggcgctg 4229

QY 4233 gccgaagagaaatccataaactttaacgagacaggtatgtccttggcagctgcacac 4292
Db 4230 gccgaagagaaatccataaactttaacgagacaggtatgtccttggcagctgcacac 4289
QY 4293 gtacacggggcactataaacaaggggtgtgttacttccagccggagggagaaataag 4352
Db 4290 gtacacggggcactataaacaaggggtgtgttacttccagccggagggagaaataag 4349
QY 4353 aacaaagagaaatctcaagcctttagaagccttacttgcacaaagagctagctatc 4412
Db 4350 aacaaagagaaatctcaagcctttagaagccttacttgcacaaagagctagctatc 4409
QY 4413 atacacttccgtgagacatcagaagagccaaagatctatctagagggagacagatgct 4472
Db 4410 atacacttccgtgagacatcagaagagccaaagatctatctagagggagacagatgct 4469
QY 4473 gaccgggttgcgaagagcagccagctgttgaacttgccttaataagaaagcctc 4532
Db 4470 gaccgggttgcgaagagcagccagctgttgaacttgccttaataagaaagcctc 4529
QY 4533 aaagcccgagaaacccagacagtagacacccctagaagagcttgcgaagagataaagata 4592
Db 4530 aaagcccgagaaacccagacagtagacacccctagaagagcttgcgaagagataaagata 4589
QY 4593 gaccagttctctgaagatcccggaagggagcctgtataacttataatggagaaatccctg 4652
Db 4590 gaccagttctctgaagatcccggaagggagcctgtataacttataatggagaaatccctg 4649
QY 4653 ccccaagaagaaggttagaataatgtccaacagatatacgtcttaacccactagagact 4712
Db 4650 ccccaagaagaaggttagaataatgtccaacagatatacgtcttaacccactagagact 4709
QY 4713 aaacacctgaagaggtgtgcagaacatcccttacttctgtgagagctacagaggt 4772
Db 4710 aaacacctgaagaggtgtgcagaacatcccttacttctgtgagagctacagaggt 4769
QY 4773 gctgaactcgtgtgtcaaacatctgtgtgacctccagctgtgtaatgtlaacttccaga 4832
Db 4770 gctgaactcgtgtgtcaaacatctgtgtgacctccagctgtgtaatgtlaacttccaga 4829
QY 4833 atacctcgaagaaagagactaagggagagccacccagcgtcgaactggagagtgagctc 4892
Db 4830 atacctcgaagaaagagactaagggagagccacccagcgtcgaactggagagtgagctc 4889
QY 4893 actgaggttaagcccggtcaaaatacgaacaaatatactatgttttggtagacacctt 4952
Db 4890 actgaggttaagcccggtcaaaatacgaacaaatatactatgttttggtagacacctt 4949
QY 4953 tcaagatgtgtgaagagcttactcactaagaagaagagcttcaacccgtgtgtgtcaaaaa 5009
Db 4950 tcaagatgtgtgaagagcttactcactaagaagaagagcttcaacccgtgtgtgtcaaaaa 5009
QY 5013 atactgtgaagaaatcttccaaagatttgaatacttaagtaalaaggttcaagaatgt 5072
Db 5010 atactgtgaagaaatcttccaaagatttgaatacttaagtaalaaggttcaagaatgt 5069
QY 5073 ccagcttctgttgcgaagtagtcaagagagctgtgcaagatattgtggatgttgaa 5132
Db 5070 ccagcttctgttgcgaagtagtcaagagagctgtgcaagatattgtggatgttgaa 5129
QY 5133 ctgcatctgtcatatacagaccccaagctcagagacaggttagagagagtaagaataccatt 5192
Db 5130 ctgcatctgtcatatacagaccccaagctcagagacaggttagagagagtaagaataccatt 5189
QY 5193 aaagagacccctac 5206
Db 5190 aaagagacccctac 5203

RESULT 14
AAC67022
ID AAC67022 standard; DNA: 4918 BP.


```

Db 3656 agttagcacaagaagctgagctcatggtccctcaagcgaagcttgcgctgcccgaagg 3715
OY 4242 aaatccataaacaatttlaacgacagcaggatgcttgcgactgcacacgtfacaagg 4301
Db 3716 aaatccataaacaatttlaacgacagcaggatgcttgcgactgcacacgtfacaagg 3775
OY 4302 gccatcataaacaagaagggttgccttaacctagcaggagggaataaagaacaagaag 4361
Db 3776 gccatcataaagaagggttgccttaacctagcaggagggaataaagaacaagaag 3835
OY 4362 gaaattctaagcctatagaagccttacatttgcacaaaaggctagctattataactgt 4421
Db 3836 gaaattctaagcctatagaagccttacatttgcacaaaaggctagctattataactgt 3895
OY 4422 cctggacatcagaagaagccaaagatc 4446
Db 3896 cctggacatcagaagaagccaaagatc 3920

```

Search completed: February 24, 2002, 07:42:18
 Job time: 16843 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 03:05:29 ; Search time 624.99 Seconds
(without alignments)
24.691 Million cell updates/sec

Title: US-09-171-553B-14

Perfect score: 18
Sequence: 1 ccacagtcgtacacacacg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18	100.0	633	22	PERV-MSL LTR SEQ I
C 2	18	100.0	704	22	IamAI LTR SEQ ID N
C 3	18	100.0	3482	20	Pig endogenous ret
C 4	18	100.0	4918	22	PERV env protein c
C 5	18	100.0	6076	22	PERV env protein c
C 6	18	100.0	6076	22	PERV env protein c
C 7	18	100.0	7333	22	PERV env protein c
C 8	18	100.0	7362	22	Defective retrovir
C 9	18	100.0	7393	18	PERV env protein c
C 10	18	100.0	7873	22	Porcine retrovirus
C 11	18	100.0	7892	18	PERV env protein c
					Miniature swine re

C 12	18	100.0	8060	18	AAT74811	Porcine retrovirus
C 13	18	100.0	8060	22	AAF77725	Tsukuba-1 cDNA. P
C 14	18	100.0	8132	22	AAF77727	Nucleotide sequenc
C 15	18	100.0	8196	18	AAV09699	Porcine retrovirus
C 16	18	100.0	8209	18	AAV09700	Porcine retrovirus
C 17	17	94.4	20	18	AAT74869	Porcine retrovirus
C 18	17	94.4	20	22	AAF77785	PCR primer #58. U
C 19	17	94.4	28	20	AAV82760	PCR primer PL148 u
C 20	15.4	85.6	6511	16	AAQ95493	Human Cdn-2 DNA.
C 21	14.8	82.2	285	22	AAH0831	Human tumour assoc
C 22	14.8	82.2	752	22	AAH04528	Human cDNA clone (
C 23	14.8	82.2	935	21	AAV09918	Fusarium venenatum
C 24	14.8	82.2	4299	19	AAV03310	Pyrolobus fumarius
C 25	14.4	80.0	3178	20	AAV99865	Human secreted pro
C 26	14.4	80.0	15894	19	AAV18265	Human secreted pro
C 27	14.4	80.0	15894	20	AAZ22902	Measles virus 1977
C 28	14.4	80.0	32249	22	AAI63718	Nucleotide sequenc
C 29	13.8	76.7	130	21	AAC14197	Human kidney relat
C 30	13.8	76.7	131	21	AAC29646	Human secreted pro
C 31	13.8	76.7	155	21	AAC13905	Human secreted pro
C 32	13.8	76.7	178	21	AAC13629	Human secreted pro
C 33	13.8	76.7	198	21	AAC23537	Human secreted pro
C 34	13.8	76.7	300	20	AAZ13103	Human gene express
C 35	13.8	76.7	300	20	AAV98595	Human cancer cell
C 36	13.8	76.7	318	21	AAC28819	Human secreted pro
C 37	13.8	76.7	344	21	AAC32109	Human secreted pro
C 38	13.8	76.7	391	22	AAF66266	Novel human polynu
C 39	13.8	76.7	417	21	AAC24025	Human secreted pro
C 40	13.8	76.7	432	22	AAH13493	Human cDNA clone (
C 41	13.8	76.7	530	21	AAZ49231	Human hydrolase ho
C 42	13.8	76.7	537	22	AAH10051	Human cDNA clone (
C 43	13.8	76.7	583	22	AAH10021	Human cDNA clone (
C 44	13.8	76.7	585	22	AAH10314	Human cDNA clone (
C 45	13.8	76.7	739	22	AAH08480	Human cDNA clone (

ALIGNMENTS

RESULT 1
AAC67007/c
ID AAC67007 standard; DNA; 633 BP.
XX AAC67007;
XX
XX 27-MAR-2001 (first entry)
XX
XX PERV-MSL LTR SEQ ID NO: 7.
XX
XX Xenotransplantation; infectious agent; vaccine; ds.
XX Porcine endogenous retrovirus.
XX
XX WO200071726-A1.
XX
XX PD 30-NOV-2000.
XX
XX 24-MAY-2000; 2000WO-US14296.
XX
XX 24-MAY-1999; 99US-0135631.
XX
XX (MAYO-) MAYO MEDICAL VENTURES.
XX
XX Federspiel MJ;
XX
XX WPI; 2001-032041/04.
XX
XX Inhibiting or preventing infectious agent transmission in mammalian
XX transplant recipients, by introducing recombinant DNA comprising DNA
XX encoding extracellular proteins of the agent into donor cells, such as
XX swine cells -
XX
XX Disclosure; Page 98; 144pp; English.

XX The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
XX Sequence 633 BP; 170 A; 125 C; 147 G; 191 T; 0 other;
SQ

Query Match 100.0%; Score 18; DB 22; Length 633;
Best Local Similarity 100.0%; Pred. No. 2.4; 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0

QY 1 ccacagtcgtacaccacg 18
|||||
DB 497 CCACAGTCGTACACCACG 480

RESULT 2
AAC67006/c
ID AAC67006 standard; DNA; 704 BP.
XX
AC AAC67006;
XX
DT 27-MAR-2001 (first entry)
XX
DE IamA1 LTR SEQ ID NO: 6.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Unidentified.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
XX 24-MAY-2000; 2000WO-US14296.
PF
XX 24-MAY-1999; 99US-0135631.
PR
XX (MAYO-) MAYO MEDICAL VENTURES.
PA
XX Federspiel MJ;
PI
XX WPI; 2001-032041/04.
DR
XX Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
XX Disclosure; Page 97-98; 144pp; English.
PS
XX The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
XX Sequence 704 BP; 187 A; 162 C; 158 G; 197 T; 0 other;
SQ

Query Match 100.0%; Score 18; DB 22; Length 704;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccacg 18
|||||
DB 568 CCACAGTCGTACACCACG 551

RESULT 3

AAV82749/c
ID AAV82749 standard; DNA; 3482 BP.
XX
AC AAV82749;
XX
DT 25-FEB-1999 (first entry)
XX
DE Pig endogenous retrovirus (PERV)-B envelope (env) gene region.
XX
KW Pig endogenous retrovirus; PERV-A; envelope protein; PERV-B; subtype;
KW probe; primer; detection; retrovirus; human tissue; xenotransplant;
KW primary porcine tissue; human cell line; porcine cell line; ss.
XX
OS Pig endogenous retrovirus.
XX
XX Key Location/Qualifiers
FH 911..2884
FT CDS
FT /*tag- a
FT /product= envelope_protein
XX
XX WO9853104-A2.
XX
XX 26-NOV-1998.
XX
XX 18-MAY-1998; 98WO-GB01428.
XX
XX 16-MAY-1997; 97GB-0010154.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
XX
XX Stoye JP, Weiss RA;
XX
XX WPI; 1999-045324/04.
DR
XX P-PSDB; AAW85452.
XX
XX Newly isolated nucleic acid probe capable of hybridising to either
PT the PERV-A or PERV-B env gene - useful in the detection of
PT retroviruses, and their subtypes, in a sample of porcine/human
PT tissue.
XX
XX Claim 3; Page 21-23; 36pp; English.
XX
XX The present sequence encodes a pig endogenous retrovirus (PERV)-B
CC envelope protein. PERV exists in two different subtypes, PERV-A and
CC PERV-B. The differences are reflected in sequence divergence in the
CC envelope genes. Probes and primers can be derived from the envelope
CC (env) genes of PERV-A and PERV-B. The probes and primers are used in
CC a method to detect retroviruses in a sample of porcine/human tissue,
CC particularly primary porcine tissue and human cell lines that have been
CC cultivated in the presence of a porcine cell line, or human tissue from
CC a patient with a xenotransplant. Subtype of PERV in a sample containing
CC one of the PERV env genes can also be determined.
XX
XX Sequence 3482 BP; 927 A; 854 C; 867 G; 834 T; 0 other;
SQ

Query Match 100.0%; Score 18; DB 20; Length 3482;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccacg 18
|||||
DB 3439 CCACAGTCGTACACCACG 3422

RESULT 4
AAC67022/c
ID AAC67022 standard; DNA; 4918 BP.
XX
AC AAC67022;
XX
XX 27-MAR-2001 (first entry)
DT
XX

DE PERV env protein coding sequence SEQ ID NO: 22.
 XX Xenotransplantation; infectious agent; vaccine; ds.
 XX
 OS Porcine endogenous retrovirus.
 XX
 FN WO200071726-A1.
 XX
 PD 30-NOV-2000.
 XX
 PF 24-MAY-2000; 2000WO-US14296.
 XX
 PR 24-MAY-1999; 99US-0135631.
 XX
 PA (MAYO-) MAYO MEDICAL VENTURES.
 XX
 PI Federspiel MJ;
 XX
 DR WPI; 2001-032041/04.
 XX
 XX Inhibiting or preventing infectious agent transmission in mammalian
 XX transplant recipients, by introducing recombinant DNA comprising DNA
 XX encoding extracellular proteins of the agent into donor cells, such as
 XX swine cells -
 PS Claim 16; Page 109-111; 144pp; English.
 XX
 CC The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 XX
 SQ Sequence 4918 BP; 1382 A; 1173 C; 1183 G; 1180 T; 0 other;
 XX
 Query Match 100.0%; Score 18; DB 22; Length 4918;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ccacagtcgtaccacag 18
 DB 4781 CCACAGTCGTACACCACG 4764
 ACACG7021/C
 ID AAC67021 standard; DNA; 6076 BP.
 XX
 AC AAC67021;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE PERV env protein coding sequence SEQ ID NO: 21.
 XX
 KW Xenotransplantation; infectious agent; vaccine; ds.
 XX
 OS Porcine endogenous retrovirus.
 XX
 FN WO200071726-A1.
 XX
 PD 30-NOV-2000.
 XX
 PF 24-MAY-2000; 2000WO-US14296.
 XX
 PR 24-MAY-1999; 99US-0135631.
 XX
 PA (MAYO-) MAYO MEDICAL VENTURES.
 XX
 PI Federspiel MJ;
 XX
 DR WPI; 2001-032041/04.
 XX

PT Inhibiting or preventing infectious agent transmission in mammalian
 PT transplant recipients, by introducing recombinant DNA comprising DNA
 PT encoding extracellular proteins of the agent into donor cells, such as
 PT swine cells -
 PS Claim 16; Page 107-109; 144pp; English.
 XX
 CC The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 XX
 SQ Sequence 6076 BP; 1613 A; 1512 C; 1609 G; 1342 T; 0 other;
 XX
 Query Match 100.0%; Score 18; DB 22; Length 6076;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ccacagtcgtaccacag 18
 DB 617 CCACAGTCGTACACCACG 600
 AAC67032/C
 ID AAC67032 standard; DNA; 6076 BP.
 XX
 AC AAC67032;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE PERV env protein coding sequence SEQ ID NO: 32.
 XX
 KW Xenotransplantation; infectious agent; vaccine; ds.
 XX
 OS Porcine endogenous retrovirus.
 XX
 FN WO200071726-A1.
 XX
 PD 30-NOV-2000.
 XX
 PF 24-MAY-2000; 2000WO-US14296.
 XX
 PR 24-MAY-1999; 99US-0135631.
 XX
 PA (MAYO-) MAYO MEDICAL VENTURES.
 XX
 PI Federspiel MJ;
 XX
 DR WPI; 2001-032041/04.
 XX
 CC The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 XX
 SQ Sequence 6076 BP; 1619 A; 1505 C; 1598 G; 1354 T; 0 other;
 XX
 Query Match 100.0%; Score 18; DB 22; Length 6076;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
 XX ||||||||||||||||
 DB 617 CCACAGTCGTACACACG 600

RESULT 7
 AAF7726/C
 ID AAF77726 standard; DNA: 7333 BP.

AC AAF77726;

DT 23-MAY-2001 (first entry)

DE Defective retroviral genome isolated from PK-15 cell line.

KW Retrovirus; graft transplantation; xenotransplantation; PK-15 cell line;

OS ss.

XX Unidentified.

PN US6190861-B1.

PD 20-FEB-2001.

PF 13-DEC-1996; 96US-0766528.

PR 14-DEC-1995; 95US-0572645.

PA (GENO) GEN HOSPITAL CORP.

PI Fishman JA;

XX WPI: 2001-256211/26.

DR P-PSDB: AAB73282, AAB73283, AAB73284.

XX Assessing risk of endogenous retroviruses in clinical practice and in

PT xenotransplantation, comprises using probe sequences derived from swine

XX or miniature swine retroviral genome -

PS Claim 1; Fig 2; 127pp; English.

XX The present invention relates to a method for screening a cell or tissue

CC for the presence or expression of a retrovirus (RV), comprising

CC contacting a target nucleic acid from the cell or tissue with a second

CC nucleic acid from the present invention (e.g. the present sequence or a

CC fragment thereof). The method is useful for RV detection and to assess

CC of donors with active replication of known viruses. Inactive proviruses

CC can be detected and inactivated, allowing identification and elimination

CC of potential human pathogens derived from swine in a manner not possible

CC in the outbred human organ donor population and its important to the

CC development of human xenotransplantation.

XX Sequence 7333 BP; 1984 A; 1788 C; 1885 G; 1676 T; 0 other;

SO

Query Match 100.0%; Score 18; DB 22; Length 7333;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
 XX ||||||||||||||||
 DB 7279 CCACAGTCGTACACACG 7262

RESULT 8

AAC67019/C

XX AAC67019 standard; DNA: 7362 BP.

DT 27-MAR-2001 (first entry)

XX

DE PERV env protein coding sequence SEQ ID NO: 19.

XX Xenotransplantation; infectious agent; vaccine; ds.

KW Porcine endogenous retrovirus.

OS WO200071726-A1.

XX 30-NOV-2000.

XX 24-MAY-2000; 2000WO-US14296.

XX 24-MAY-1999; 99US-0135631.

XX (MAYO-) MAYO MEDICAL VENTURES.

XX Federspiel MJ;

XX WPI: 2001-032041/04.

XX Inhibiting or preventing infectious agent transmission in mammalian

XX transplant recipients, by introducing recombinant DNA comprising DNA

XX encoding extracellular proteins of the agent into donor cells, such as

XX swine cells -

XX Claim 16; Page 101-104; 144pp; English.

XX The present invention provides a method to prevent the transmission of

XX infectious agents during xenotransplantation. This involves introducing

XX to donor swine cells a recombinant DNA encoding a peptide fragment from

XX the infectious agent, and then introducing these cells into the

XX transplant recipient.

XX Sequence 7362 BP; 1997 A; 1821 C; 1881 G; 1663 T; 0 other;

SO

Query Match 100.0%; Score 18; DB 22; Length 7362;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
 XX ||||||||||||||||
 DB 682 CCACAGTCGTACACACG 665

RESULT 9

AAT74883/C

XX AAT74883 standard; cDNA: 7393 BP.

AC AAT74883;

DT 09-FEB-1998 (first entry)

XX Porcine retrovirus cDNA (defective).

XX Retrovirus; porcine; GAG protein; POL protein; ENV protein;

XX xenotransplantation; infectious; provirus; organ transplant; donor;

XX activated virus; PCR; ss.

XX Porcine retrovirus.

XX Key Location/Qualifiers

FT CDS 598..2172

FT mat_peptide 598..2169

FT CDS 2320..4737

FT mat_peptide 2320..4737

FT mat_peptide 2320..3522

FT /tag= d

FT /note= "putative GAG protein"

FT /note= "putative POL coding region (partial) as described in the specification"


```

FT mat_peptide /note="putative POL protein (partial)"
FT 3516..4328 /*tag= e
FT /note="putative POL protein (partial)"
FT mat_peptide 4332..4748
FT /*tag= f
FT CDS 4738..6725
FT /*tag= g
FT /note="putative ENV coding region (partial) as
FT mat_peptide 4752..6722
FT /*tag= h
FT /note="ENV protein (partial)"
FT W09721836-A1.
FT 19-JUN-1997.
FT 13-DEC-1996; 96WO-US19680.
FT 14-DEC-1995; 95US-0572645.
FT (GEHO ) GEN HOSPITAL CORP.
FT Fishman JA;
FT WPI; 1997-332804/30.
FT P-PSDB; AAW32091-W32095.
FT New nucleic acid from porcine retroviruses - used for detecting
FT viruses in transplant or other tissue and for assessing risk of
FT transmitting infection to graft recipient
FT Claim 16; Fig 2; 128bp; English.
XX
XX This cDNA sequence represents a defective purified swine retrovirus
XX found in PK-15 cells containing the putative coding regions for viral
XX GAG, POL and ENV proteins. There are a few in frame stop codons and
XX apparent frame shifts in the given coding sequence which alter features
XX of the translation. This sequence and PCR fragments generated from the
XX sequence (see AAT74812-774882) could be used to screen organs for the
XX presence of porcine retroviruses prior to xenotransplantation.
XX Transplantation can increase the likelihood of retroviral activation if
XX intact and infectious proviruses are present. The porcine retroviral
XX sequence can be used to generate probes to determine the level (e.g.
XX copy number) of intact (i.e. potentially replicating) porcine provirus
XX sequences in a strain of xenograft transplantation donors. It can be
XX used to detect mutations, genetic lesions or viral recombinants and to
XX determine the histological localisation of activated retroviruses. Using
XX Polymerase Chain Reaction DNA Quantitation (PQ) on blood mononuclear
XX cells, infectivity titration and susceptibility testing can be
XX performed. Ultimately animal donors without intact porcine retroviral
XX sequences or a lower copy number of viral elements could be selected.
SQ Sequence 7393 BP; 2004 A; 1801 C; 1896 G; 1692 T; 0 other;

```

```

Query Match 100.0%; Score 18; DB 18; Length 7393;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ccacagtcgtaccacg 18
DB 7339 CCACAGTGTACACG 7322

```

```

RESULT 10
AAC67023/C
ID AAC67023 standard; DNA: 7873 BP.
XX AAC67023;
XX 27-MAR-2001 (first entry)

```

```

XX PERV env protein coding sequence SEQ ID NO: 23.
DE Xenotransplantation; infectious agent; vaccine; ds.
XX Xenotransplantation; infectious agent; vaccine; ds.
XX Porcine endogenous retrovirus.
XX W0200071726-A1.
XX 30-NOV-2000.
XX 24-MAY-2000; 2000WO-US14296.
XX 24-MAY-1999; 99US-0135631.
XX (MAYO-) MAYO MEDICAL VENTURES.
XX Federspiel MJ;
XX WPI; 2001-032041/04.
XX Inhibiting or preventing infectious agent transmission in mammalian
XX transplant recipients, by introducing recombinant DNA comprising DNA
XX encoding extracellular proteins of the agent into donor cells, such as
XX swine cells -
XX Claim 16; Page 112-115; 144pp; English.
XX
XX The present invention provides a method to prevent the transmission of
XX infectious agents during xenotransplantation. This involves introducing
XX to donor swine cells a recombinant DNA encoding a peptide fragment from
XX the infectious agent, and then introducing these cells into the
XX transplant recipient.
XX
SQ Sequence 7873 BP; 2200 A; 1913 C; 1962 G; 1798 T; 0 other;

```

```

Query Match 100.0%; Score 18; DB 22; Length 7873;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ccacagtcgtaccacg 18
DB 7708 CCACAGTGTACACG 7691

```

```

RESULT 11
AAT74884/C
ID AAT74884 standard; cDNA: 7892 BP.
XX AAT74884;

```

```

DT 09-FEB-1998 (first entry)

```

```

XX Miniature swine retrovirus cDNA.

```

```

XX Retrovirus: porcine; GAG protein; POL protein; ENV protein;
XX xenotransplantation; infectious; provirus; organ transplant; donor;
XX activated virus; PCR; ss.

```

```

XX Porcine retrovirus.

```

```

XX Key Location/Qualifiers
XX CDS 585..2159

```

```

XX mat_peptide

```

```

XX /*tag= a
XX 585..2156
XX /*tag= b

```

```

XX /note="putative GAG protein"

```

```

XX CDS 2307..5744
XX /*tag= c

```

```

XX mat_peptide 2307..5741
XX /*tag= d

```

```

XX /note="putative POL protein"

```

```

FT CDS 5620..7536
FT /*tag= e
FT /*note= "putative ENV protein"
FT mat_peptide
FT /*tag= f
FT /*note= "putative ENV protein"
FT
PN WO9721836-A1.
PD 19-JUN-1997.
PF 13-DEC-1996; 96MO-US19680.
PR 14-DEC-1995; 95US-0572645.
PA (GENO ) GEN HOSPITAL CORP.
PI Fishman JA;
DR WPI; 1997-332804/30.
DR P-PSDB; AAM32096-W32098.
XX
XX New nucleic acid from porcine retroviruses - used for detecting
PT viruses in transplant or other tissue and for assessing risk of
PT transmitting infection to graft recipient
XX
XX Claim 22; Fig 3; 128pp; English.
XX
XX This CDNA sequence represents a porcine retrovirus from miniature swine
CC containing the putative coding regions for viral GAG, POL and ENV
CC proteins. This sequence and PCR fragments generated from it
CC (see AAT74812-T74882) can be used to screen organs for the presence of
CC porcine retroviruses prior to xenotransplantation. Transplantation can
CC increase the likelihood of retroviral activation if intact and
CC infectious proviruses are present. The porcine retroviral sequence can be
CC used to generate probes to determine the level (e.g. copy number) of
CC intact (i.e. potentially replicating) porcine provirus sequences in a
CC strain of xenograft transplantation donors. It can be used to detect
CC mutations, genetic lesions or viral recombinants and also to determine
CC the histological localization of activated retrovirus. Using Polymerase
CC Chain Reaction DNA Quantitation (PQ) on blood mononuclear cells,
CC infectivity titration and susceptibility testing can be performed.
CC Ultimately animal donors without intact porcine retroviral sequences or a
CC lower copy number of viral elements could be selected.
XX
XX Sequence 7892 BP; 2191 A; 1915 C; 1980 G; 1806 T; 0 other;
SQ
Query Match 100.0%; Score 18; DB 18; Length 7892;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ccacagtcgtacaccag 18
DB 19 CCACAGTCGTACACCAG 2
RESULT 12
AAT74811/C
ID AAT74811 standard; cDNA; 8060 BP.
XX
XX AAT74811;
XX
XX 11-FEB-1998 (first entry)
XX
XX Porcine retrovirus Tsukuba-1 cDNA.
XX
XX Retrovirus; porcine; GAG protein; POL protein; ENV protein;
KW xenotransplantation; infections; provirus; organ transplant; donor;
XX activated virus; Tsukuba-1; PCR; ss.
XX
XX Porcine retrovirus.
XX
XX Key Location/Qualifiers

```

```

FT CDS 86..2002
FT /*tag= a
FT /*note= "putative ENV protein"
FT CDS 3112..4686
FT /*tag= b
FT /*note= "putative GAG protein"
FT CDS 4871..8060
FT /*tag= c
FT /*note= "putative POL protein (partial)"
FT
PN WO9721836-A1.
PD 19-JUN-1997.
PF 13-DEC-1996; 96MO-US19680.
PR 14-DEC-1995; 95US-0572645.
PA (GENO ) GEN HOSPITAL CORP.
PI Fishman JA;
DR WPI; 1997-332804/30.
XX
XX New nucleic acid from porcine retroviruses - used for detecting
PT viruses in transplant or other tissue and for assessing risk of
PT transmitting infection to graft recipient
XX
XX Claim 1; Fig 1; 128pp; English.
XX
XX This sequence represents the purified porcine retroviral cDNA
CC sequence of Tsukuba-1 and contains the putative coding regions for viral
CC proteins GAG, POL and ENV. This sequence and PCR fragments generated
CC from the sequence (see AAT74812-T74882) could be used to screen organs
CC for porcine retroviruses prior to xenotransplantation. Transplantation
CC can increase the likelihood of retroviral activation if intact and
CC infectious proviruses are present. The porcine retroviral sequence can be
CC used to generate probes to determine the level (e.g. copy number) of
CC intact (i.e. potentially replicating) porcine provirus sequences in a
CC strain of xenograft transplantation donors. It can be used to detect
CC mutations, genetic lesions or viral recombinants and to determine the
CC histological localization of activated retrovirus. Using Polymerase Chain
CC Reaction DNA Quantitation (PQ) on blood mononuclear cells, infectivity
CC titration and susceptibility testing can be performed. Ultimately animal
CC donors without intact porcine retroviral sequences or with a lower copy
CC number of viral elements could be selected.
XX
XX Sequence 8060 BP; 2233 A; 1959 C; 2012 G; 1856 T; 0 other;
SQ
Query Match 100.0%; Score 18; DB 18; Length 8060;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ccacagtcgtacaccag 18
DB 2546 CCACAGTCGTACACCAG 2529
RESULT 13
AAF77725/C
ID AAF77725 standard; cDNA; 8060 BP.
XX
XX AAF77725;
XX
XX 23-MAY-2001 (first entry)
XX
XX Tsukuba-1 cDNA.
XX
XX Tsukuba-1; retrovirus; graft transplantation; xenotransplantation; ss.
XX
XX Porcine retrovirus.
XX
XX Key Location/Qualifiers

```

PN US6190861-B1.
PD 20-FEB-2001.
XX
XX
PF 13-DEC-1996; 96US-0766528.
XX
PR 14-DEC-1995; 95US-0572645.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Fishman JA;
DR WPI; 2001-256211/26.
XX
PT Assessing risk of endogenous retroviruses in clinical practice and in
PT xenotransplantation, comprises using probe sequences derived from swine
or miniature swine retroviral genome -
PS Claim 1; Fig 1; 127pp; English.

CC The present invention relates to a method for screening a cell or tissue
CC for the presence or expression of a retrovirus (RV), comprising
CC contacting a target nucleic acid from the cell or tissue with a second
CC nucleic acid from the present invention (e.g. the present sequence or a
CC fragment thereof). The method is useful for RV detection and to assess
CC of donors with active replication of known viruses. Inactive proviruses
CC can be detected and inactivated, allowing identification and elimination
CC of potential human pathogens derived from swine in a manner not possible
CC in the outbred human organ donor population and is important to the
CC development of human xenotransplantation.

Sequence 8060 BP; 2233 A; 1959 C; 2012 G; 1856 T; 0 other;

Query Match	100.0%	Score 18;	DB 22;	Length 8060;
Best Local Similarity	100.0%	Pred. No. 2.7;		
Matches 18; Conservative	0;	Mismatches	0;	Gaps 0;
QY	1	ccccagtcgtacacacag	18	
Db	2546	CCACAGTCGTACCCACG	2529	

```

RESULT 14
AAE77727/c
ID AAE77727 standard; DNA; 8132 BP.
XX
XX
AC AAE77727;
XX
XX
DT 23-MAY-2001 (first entry)
XX
XX
DE Nucleotide sequence of a retrovirus found in miniature swine.
XX
XX
KW Retrovirus; graft transplantation; xenotransplantation; miniature swine
XX
XX
ss.
XX
OS Unidentified.
XX
XX
PN US6190861-B1.
XX
XX
PD 20-FEB-2001.
XX
XX
PF 13-DEC-1996; 96US-0766528.
XX
XX
PR 14-DEC-1995; 95US-0572645.
XX
XX
PA (GENO ) GEN HOSPITAL CORP.
XX
XX
PI Fishman JA;
XX
XX
WPI: 2001-256211/26
DR P-PSDB: AAE73285, AAB73286, AAB73287.
DR

```

XX Assessing risk of endogenous retroviruses in clinical practice and in
PT xenotransplantation, comprises using probe sequences derived from swine
PT or miniature swine retroviral genome -
XX
XX
PS Claim 1; Fig 3; 127pp; English.
PS
CC The present invention relates to a method for screening a cell or tissue
CC for the presence or expression of a retrovirus (RV), comprising
CC contacting a target nucleic acid from the cell or tissue with a second
CC nucleic acid from the present invention (e.g. the present sequence or a
CC fragment thereof). The method is useful for RV detection and to assess
CC graft transplantation risk. Screening of animals allows the elimination
CC of donors with active replication of known viruses. Inactive proviruses
CC can be detected and inactivated, allowing identification and elimination
CC of potential human pathogens derived from swine in a manner not possible
CC in the outbred human organ donor population and is important to the
CC development of human xenotransplantation.
XX
XX
XX Sequence 8132 BP; 2248 A; 1977 C; 2037 G; 1870 T; 0 other;
50

	Query Match	100.0%;	Score 18;	DB 22;	Length 8132;
	Best Local Similarity	100.0%;	Pred. NO. 2.7;		
Matches	18;	Conservative	0;	Mismatches	0; Gaps 0;
OY	1 ccacagtcgttaaccacg	18			
DG	8080 CCACAGTCGTACCACCG	8063			

RESULT	15
AAV09699/c	
ID	AAV09699 standard; DNA; 8196 BP.
XX	
AC	AAV09699;
XX	
DT	19-MAY-1998 (first entry)
XX	
DE	Porcine retrovirus DNA encoding, GAG, POL and ENV.
XX	
KW	Porcine retrovirus; poEV; POL protein; ENV protein; GAG protein;
XX	
XX	vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds
XX	
XX	Porcine retrovirus.
XX	
PH	Key
FT	Location/Qualifiers
FT	576..2126
FT	/*tag= a
FT	/product= GAG protein
FT	/note= "villon core polypeptide"
FT	2143..5733
FT	/*tag= b
FT	/product= POL protein
FT	/note= "polymerase peptide sequence as given in
FT	5606..7576
FT	/*tag= c
FT	/product= ENV protein
FT	/note= "envelope protein"
XX	
PN	MO9740167-A1.
XX	
PD	30-OCT-1997.
XX	
PF	18-APR-1997; 97WO-GH01087.
XX	
PR	10-FEB-1997; 97GB-0002668.
PR	19-APR-1996; 96GB-0008164.
XX	
PA	(IMUT-) IMUTRAN LTD.
PA	(OONE-) O-ONE BIOTECH LTD.
XX	

PI Galbraith DN, Haworth C, Lees GM, Smith KT;
 XX WPI, 1997-535851/49.

DR Polynucleotide encoding porcine retrovirus expression product -
 XX useful to develop products for use in vaccines, diagnosis and
 PT xeno-transplantation
 PS
 XX

PS Claim 4; Fig 2; 69pp; English.

XX This DNA sequence encodes the porcine retrovirus (PoEV) virion core
 CC polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins. These
 CC proteins can be used to develop viral vaccines, antisense nucleic acids,
 CC ribozymes and other antiviral agents. They can also be used in
 CC xeno-transplantation technology and as diagnostic tools.

XX Sequence 8196 BP; 2165 A; 2061 C; 2147 G; 1820 T; 3 other;

Query Match 100.0%; Score 18; DB 18; Length 8196;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccacg 18

|||||

DB 8130 CCACAGTCGTACACCACG 8113

Search completed: February 24, 2002, 03:05:32
 Job time: 19498 sec

Mon Feb 25 07:44:48 2002

us-09-171-553b-1.feb2201i.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 02:56:31 ; Search time 197.25 Seconds
(without alignments)
3811.949 Million cell updates/sec

Title: US-09-171-553b-1

Perfect score: 3320
Sequence: 1 gaattcgagcgccgctgcac.....aaaaaaaaaaaaaaaa 3320

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCRNUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	1091	32.9	7333	4 US-08-766-528-2	Sequence 2, Appli
2	786	23.7	2462	3 US-09-111-085-1	Sequence 1, Appli
3	704	21.2	8132	4 US-08-766-528-3	Sequence 3, Appli
4	375	11.3	8060	4 US-08-766-528-1	Sequence 1, Appli
5	321	9.7	3482	3 US-09-111-085-3	Sequence 3, Appli
6	265	8.0	445	4 US-09-376-781-7	Sequence 7, Appli
7	35	1.1	8655	3 US-09-075-272-1	Sequence 7, Appli
8	30	0.9	1279	4 US-09-277-716-31	Sequence 31, Appli
9	30	0.9	735	1 US-08-798-000-3	Sequence 1, Appli
10	29	0.9	1198	3 US-09-248-335-27	Sequence 27, Appli
11	29	0.9	1234	1 US-08-798-000-1	Sequence 1, Appli
12	29	0.9	2338	4 US-09-232-200-66	Sequence 66, Appli
13	29	0.9	2338	4 US-09-232-197-66	Sequence 66, Appli
14	29	0.9	7308	4 US-09-011-745-4	Sequence 4, Appli
15	29	0.9	7308	4 US-09-011-745-3	Sequence 2, Appli
16	29	0.9	7616	4 US-09-011-745-2	Sequence 13, Appli
17	29	0.9	8202	1 US-08-258-420-13	Sequence 1, Appli
18	29	0.9	8332	1 US-08-850-961-1	Sequence 5, Appli
19	29	0.9	140	1 US-08-628-417-5	Sequence 34, Appli
20	28	0.8	216	1 US-08-686-878A-34	Sequence 6, Appli
21	28	0.8	246	1 US-08-628-417-6	Sequence 24, Appli
22	28	0.8	330	4 US-09-078-294-24	Sequence 278, Appli
23	28	0.8	595	4 US-09-385-982-278	Sequence 18, Appli
24	28	0.8	698	4 US-09-376-781-18	Sequence 1, Appli
25	28	0.8	719	1 US-08-375-346A-1	Sequence 1, Appli
26	28	0.8	719	2 US-08-467-123B-1	Sequence 1, Appli
27	28	0.8	719	2 US-08-467-123B-1	Sequence 1, Appli

28	0.8	949	5 PCT-US91-08254-1	Sequence 1, Appli
29	0.8	949	5 PCT-US91-08254-2	Sequence 2, Appli
30	0.8	1132	3 US-08-894-731-3	Sequence 3, Appli
31	0.8	1578	4 US-09-416-050A-1	Sequence 1, Appli
32	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
33	0.8	1578	4 US-09-665-309-1	Sequence 1, Appli
34	0.8	1578	4 US-09-661-569-1	Sequence 7, Appli
35	0.8	1964	3 US-08-468-856B-7	Sequence 12, Appli
36	0.8	1964	3 US-08-468-859A-7	Sequence 25, Appli
37	0.8	2000	4 US-09-376-781-25	Sequence 30, Appli
38	0.8	2000	4 US-09-376-781-30	Sequence 14, Appli
39	0.8	3291	4 US-09-318-448-12	Sequence 8, Appli
40	0.8	43	2 US-08-686-599A-14	Sequence 3, Appli
41	0.8	104	2 US-08-803-899-8	Sequence 3, Appli
42	0.8	153	4 US-09-244-794A-3	Sequence 17, Appli
43	0.8	153	4 US-09-247-190-3	Sequence 17, Appli
44	0.8	159	4 US-09-244-794A-17	Sequence 17, Appli
45	0.8	159	4 US-09-247-190-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ. ID NO.: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-2

Query Match 32.9%, Score 1091; DB 4; Length 7333;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
1364 aaagggggttggcttcacgaggaaggaataaagaaaggaattcgaagc 1423
|||||

Db	3517	AAAGGGGGTTGGCTTACTCAGCAGGAGGGGAAATTAAGAACAAGAGAAATTTCAAGCC	3576
Qy	1424	tattagaagccttaacatttgcacaaaaggctagctctlaaacctgctctgtaactaga	1483
Db	3577	TATTAGAAAGCCTTACATTGGCCAAAAGGCTAGCTATTATPACACTGTCTTGACATCGA	3636
Qy	1484	aagccaagaatctcatatcttagaggaaacagatgctgtaaccggtltgcacagcagag	1543
Db	3637	AAOCCAAAGATCTCATATCTACAGGGGAAACAAATGCTGACCGGTTGCCAAGCAGGAG	3696
Qy	1544	cccaagcgtgttaaccttctgcctataatagaagcccaagccccaagaaacccaagaac	1603
Db	3697	CCCAAGCCTGTAAACCTTCTGCTATTATAGAAACGCCCAAGGCCCAAGACCCAGACAG	3756
Qy	1604	agtaaaccttagaagaatgcygaagaagataaaaaagataagaccagttctctgagatccg	1663
Db	3757	AGTAAACCCCTAAGAAAGCTGGCAAGAGATAAAAAAGATAGACCACTTCTCTGAGACTCGG	3816
Qy	1664	aggggaactgctatacctcatactgaaatctggaagaaatccctgcgccccaagaagaggt	1723
Db	3817	AAGGAGCCTGCTATCTACTCATATGGGAAGAAATCTCTCCCCACAAAGAGGGTGAAT	3876
Qy	1724	atgtcaacaagaatatactgttaacccaacttagaactaaacactgacagcagltgtca	1783
Db	3877	ATGTCACAACATATACATGCTGTAACCCACCTAGGAACATAAACACCTGACACATTTGTCA	3936
Qy	1784	gaacatcccttatactgltcttgagagcttacagaagtggtgtgctctggtgtgtcaaat	1843
Db	3937	GAACATCCCTTATCATATGTTCTGTGAGCTACAGGAGTGGCTGACTGGTGGTCAAACTT	3996
Qy	1844	gtgtgccctgcagctggttaatgcttaatcttcccaagaatcactccaagaaagagactaa	1903
Db	3997	GTGTGCCCTGCACACTGGTTAAATGCTAATCTTTCGAAGTATACCTCCAGGAAAGACATTA	4056
Qy	1904	ggggaagcccaacccagagcgtcactggygaagtggactcaactgagttaaagccggtaat	1963
Db	4057	GGGGAACCTCAACCCAGGGGCTCATCGGGAAGTGGACTTCATGAGTAAAGCCGGCTAAT	4116
Qy	1964	acggaacaatatcatattglttltgttgacaaccttccagatgggtagaggtctatc	2023
Db	4117	ACGGAACAATATCTATTGTTGTTTGTAGAACACTTTTCAGAGTGGGTAGAGGCTTATC	4176
Qy	2024	cttctaagaagaagactcaaccgtgtgtgctcaagaataactgtgaggaatltttccaa	2083
Db	4177	CTACTAAAAAAGACTTCAACCGTGTGCTTAACAATAATCTGGAGSAAATTTTTCAA	4236
Qy	2084	gatttgaataaccttaagatgaatagggtcaagaatggtccagcttgcgttgccaggttaa	2143
Db	4237	GATTGGATTAACCTAAAGTAAATGAGGTCAAGAAATGATCCAGCTTGTGTTGCCAGGTAA	4296
Qy	2144	gtcaagagactggtccaaagatatltgggacttgatttgaacatgcatltgtgtcaagaaccc	2203
Db	4297	GTCAGGAGATGGCCAAAGATATTGGGGATTTGTAAGAAATCTGATTTGTCAATACAGACCC	4356
Qy	2204	aaagctcagagaaggttagagaggtatgaatagaaccatlaaagagacccttacaatatga	4416
Db	4357	AAAGCTCAAGACAGGTATAGAGAGATGAATTAACCATTTAAAGAGACCTTACCAATTTGA	4476
Qy	2264	ccaagagaactgcatatgaatgaatgtgtgtgtccctgaccttggcttlttaagggtga	2323
Db	4417	CCACAGAGACTGGCATTAATGATTGGATGCTCTCTCGCCCTTTGTGGCTTTTAAAGGTGA	4476
Qy	2324	ggaacacccctggagacglttgggctgtaaccctatgaattgctctcaaggggagacccccc	2383
Db	4477	GGAAACCCCTGGAGACGTTTGGGTGACCCCTCATTAATTTGGCTTACGGGGACCCCCCC	4536
Qy	2384	cgttggacagaatltgcttgcatactagtgatgtgtgtgttccagccttggctt	2443
Db	4537	CGTTGGCAGAAATGCTTTGGACATATGTCATGTCATGTCGCTTTCCAGCCTTTGTTCT	4596
Qy	2444	cttaggtccaagagcgtctcagtggtgtgaagcagcagcgtgtaagcagcttccggggagcct	2503
Db	4597	CTAAGCTCAAGGCTCTGACTGGGTGAGGACAGACCGTGGAGACGACTCCGGGAGGGCTT	4556

QY	2504	accacgagggagacttcgaagttccacatcgcttccaaattggaattcagtcatagtta	256
Db	4657	ACTCAAGGAGGAACTTCCAAAGTTCACATCGCTCCAAAGTGGAGATTAGCTATGTTA	4716
QY	2554	gaacgcacccgtgcaggaagacctgaacacctcggtggaaggaagaccttatctgactattga	2623
Db	4717	GAGCCACCGTGCAGGAACCTCGAAGCTGGTAGAAGGACCTATCTCGATTTTGA	4776
QY	2624	ccacaccaacgctgtgtgaagtcgtaaggaatcccc	2658
Db	4777	CCACACCAACGCGTGTGAAGAATGGAAGGAATCCCC	4811

```

RESULT 2
US-09-111-085-1
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 438/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2462
; TYPE: DNA
; ORGANISM: Porcine retrovirus
; US-09-111-085-1

```

Query Match	Best Local Similarity	Score	DB 3	Length
Matches 836; Conservative	99.98;	786;	0;	2462;
	0;		1;	
QY 2459	tcgagttggtgtagcgacgagcttggaacagctccggagagcctctcaagaagagact	2518		
DB 1	tcgagtcgggtgtagcgacgagcgttggaacagctccggagagcctctcaagaagagact	60		
QY 2519	tggaaattccacatcgcttccaaagtgtgagatcagtcattgttagacgcacgcgtgag	2578		
DB 61	tgcgaattccacatcgcttccaaagtgtgagatcagtcattgttagacgcacgcgtgag	120		
QY 2579	gaaacctcgagagctcgtgttgaaaggagccttactctgacttttgacacacacagcgtg	2638		
DB 121	gaaacctcgagagctcgtgttgaaaggagccttactctgacttttgacacacacagcgtg	180		
QY 2639	tgaagctcgaaagaaatccccaactgcatcattgcatccacgcttgaagccgagccacctc	2698		
DB 181	tgaagctcgaaagaaatccccaactgcatcattgcatccacgcttgaagccgagccacctc	240		
QY 2699	cggattcgggtgtggaagacgcgaagaagacttgaanaatcccttaagcttcgctccatcg	2758		
DB 241	cggattcgggtgtggaagacgcgaagaagacttgaanaatcccttaagcttcgctccatcg	300		
QY 2759	tgtttccctactctgtcaatactcctcaagttaaagttaaagcgccttgtgaacagccg	2818		
DB 301	tgtttccctactctgtcaatactcctcaagttaaagttaaagcgccttgtgaacagccg	360		
QY 2819	aactccataaaaccttactctcaacttgtttaacttcaaccctggtaaggtataat	2878		
DB 361	aactccataaaaccttactctcaacttgtttaacttcaaccctggtaaggtataat	420		
QY 2879	altaacgcactcaagggagagctcccttggggagcttggcgtcgtgaatatatgtctgc	2938		
DB 421	altaacgcactcaagggagagctcccttggggagcttggcgtcgtgaatatatgtctgc	480		

QY 2939 ctctgacgaatccctgtctcaatgacagcgaccccccgatctccgtgct 2998
Db 481 ctctgacgaatccctgtctcaatgacagcgaccccccgatctccgtgct 540
QY 2999 taagggttcaagcttgcacagagaccccccaataatgaagaatattgtgaaalccctcag 3058
Db 541 taagggttcaagcttgcacagagaccccccaataatgaagaatattgtgaaalccctcag 600
QY 3059 gattcccttgcaagcaatgagctgtacatacttctaaatgaggaattgtgaaatgtgca 3118
Db 601 gattcccttgcaagcaatgagctgtacatacttctaaatgaggaattgtgaaatgtgca 660
QY 3119 gtctctgaagcaagagatgaatgaatctcttcttgaacatccctaccagttatacga 3178
Db 661 gtctctgaagcaagagatgaatgaatctcttcttgaacatccctaccagttatacga 720
QY 3179 tttaattatgcatgagatgagatgaaagattgtgcaagcggtgtacaaaagatgtgca 3238
Db 721 tttaattatgcatgagatgagatgaaagattgtgcaagcggtgtacaaaagatgtgca 780
QY 3239 aataagcaataagctgtcattctgttagacctaattacttaaaatgaatttact 3295
Db 781 aataagcaataagctgtcattctgttagacctaattacttaaaatgaatttact 837

RESULT 3

US-08-766-528-3

Sequence 3, Application US/08766528

Patent No. 6190861

GENERAL INFORMATION:

APPLICANT: Jay A. Fishman

TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766,528

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/572,645

FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers

REGISTRATION NUMBER: 35,965

REFERENCE/DOCKET NUMBER: MGP-038CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 8132 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-766-528-3

Query Match 21.2%; Score 704; DB 4; Length 8132;
Best Local Similarity 98.7%; Pred. No. 2.8e-299;
Matches 2204; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 22 gatgccttctctgcctgagatlaacacccactagaccaacacttttgcccttgatgg 81
Db 2970 GATGCTTCTCTGCTGAGATTACACCCACAGCAACCACTTTTCTCTCAATGG 3029
QY 82 agagatccaggtacgaggaacccggagcctcactgtgacccgactgtgcccgaaggttc 141
Db 3030 AGAGATCCAGGTACGGGAAGAACCCGGCAGCTCACTGGACCCGACTGCCCCAAGGTTG 3089
QY 142 aagaactcccccacacacttttgaagagccctacacagagacctgtgccaacttcagatc 201
Db 3090 AAGAATCTCCCGACACACTTTTGAAGAGCCCTTACACAGAGACTGTGGCCAACTTCAGATG 3149
QY 202 caacacccctcaagltgacccctccacagtaagltgaaactgtcttctgcygagaccac 261
Db 3150 CAACACCTTCAGGTAGACCCCTCCAGTACGTAGATGACTGCTTGTGGGAGACAC 3209
QY 262 aaacaggaactgtctagaaggtacgaaggaactactgtcgtgaattgtctgaactaggtac 321
Db 3210 AAACAGGACTGTCTAGAGAGCAAGAGCACTACTGCTGAATGTCTGACTAGGCTTAC 3269
QY 322 agagcctctgtctgaagagagccagatttgcagagagaggtaaactatcttggtgatacgt 381
Db 3270 AGAGCCTCTGTCTAGAGAGCCCGCATTTGCAAGAGAGAGTAACTTTGGGTTCACT 3329
QY 382 ttgcgggcygagcagatgtgtacgagcagcagcaggaagaaactgtlgtccagataccg 441
Db 3330 TTGCGGGACGAGGAGATGCTGACGAGAGGACGAGAAACTGTAGTCCAGATACCG 3389
QY 442 gcccccacacagcacaagaagtgaagatttttgggagccgtgattttgcaagctg 501
Db 3390 GCCCCAAACACACAGCAAAAGAGAGATTTTGGGACACTGTGATTTTGGACACATG 3449
QY 502 tggatcccgagggtttgcagacttaagcagcccccactacccgttaaccaaagaagag 561
Db 3450 TGGATCCCGAGGTTTGGAGACTTAGACGCCCCACTTACCCGTAAACCAAGAAAGAG 3509
QY 562 ggaattctctgtggtcctcagacacacgaaggaattgtatgtactcaaaaagccctgtg 621
Db 3510 GAATTCCTGTGGGCTCTGTAGACACAGAGAGCTTTGTATGCTTCAAAAAGGCGTGTG 3569
QY 622 agcgcaactgtctctgcccctcctgaagtaactaaaccccttaacttattgtgatac 681
Db 3570 AGCGCACTCTCTGTGCGCTCCCTGAGCTAGCTAAACCTTTTACCTTTATGTGGATGAG 3629
QY 682 cgttaaggaatagccgaggaaggttlaacccaaacacttagaacatlgaggaagacctgt 741
Db 3630 CGTAGGAGATAGCGCGGAGGATTTAACCAACCTTAGACCATGAGAGAACACTGTG 3689
QY 742 gctactcttcaagaagacttgatcctgtagcagatgtgtgcccgtatgtctgaaggtc 801
Db 3690 GCTACTCTTCAAGAAGCTTCGATCTGTAGCCAGTGTGTGGCCATATGCTTGAAGGCT 3749
QY 802 atcgcaactgttgccatactggttcaagaagcgtgacaaatcttgaccltggagagaatata 861
Db 3750 ATCGCACTGTGGCCATACGTAGGTCAAGAGCGCTGACAAATTTGACTTTGGACAGAAATATA 3809
QY 862 actgtaatagcccccaatgacttgaagagaactgttcgagagcccccaagccagatgatg 921
Db 3810 ACTGTAAATAGCCCCCATGATGATGAGACAATGCTGTGGAGCCCCCAGACCATGATGATG 3869
QY 922 accaagcccgatagaccacactatcaaaagctgtcttcaacagagaggtcacttgcgt 981
Db 3870 ACCAAGCCCGCATAGCCCATATCAAAAGCTGTGTTCTCAAGAGAGGTCAGCTTGCT 3929
QY 982 ccaccagcgccttcaacccctgcaactctctgtgcttgaagagactgtatgaacagatgact 1041
Db 3930 CCACCAGCGCTTCAACCCCTGCACTCTTGTGCTTAAGAGACTGATGAACCAAGTACT 3989
QY 1042 catgatgtgcaatcaactatgatgatgagagaggtgtccgcaaggaaccttaacagata 1101
Db 3990 CATGATGTGCCATCAACTATGATGATGAGAGACTGGGTCGCGCAAGGACCTTTACAGACATA 4049
QY 1102 ccgctgactgtgagaagtgcttaactgttcaactgtacggaagcagctatgtgtgtgaggt 1161

Db 5735 AACAGGACTGCTTAGAAGGACGACGACACTACTGCTGGAATGTCGACACTAGGCTAC 5794
QY 322 agagcctctgtaagaagccagagattgcaagagagagtaacatactggggtacgt 381
Db 5795 AGAGCCTGCTGTAAGAAAGGCCAGATTGCAAGAGAGGTAACATATCTGGGGTACGT 5854
QY 382 ttgctgggagcgagcagcatgctacgagagcagcagaagaacatgctccagataccg 441
Db 5855 TTACGGGAGCGGACGACATGGCTCAGCGAGGACGAGAAACTGTATGTCAGATACCG 5914
QY 442 gcccaacacacagcccaaaagatgagagagttttgggagcagctgagatttgcaagctg 501
Db 5915 GCCCAACACACAGCCAAATATGAGAGAGTTTTGGGGACAGCTGATTTTGGAGACTG 5974
QY 502 tggatcccgagggtttgcaaccttagcagcccccactaccgcgttaaccaagaagaagg 561
Db 5975 TGGATCCCGGGGTTGGACCTTAGCAGCCCCACTACCGGCTAACCCGCTAACCAAGAAAGGG 6034
QY 562 ggaattctctgagctcctgagcaccagaagagcaltgctatcaaaaagagccctctg 621
Db 6035 GAATTCTCTGCGCTCTGAGACACAGGACATTTGATGCTATCAAAAAGGCCCTGCTG 6094
QY 622 agcgacactctctgtgcccctccctgacgtacctaacccttacccttattgtgagtag 681
Db 6095 AGCGACACTGCTGTGGCCCTCCCTGACGTAACTAAACCTTTACCTTTATGTGATGAG 6154
QY 682 cgttaaggagtagccgagagagttttaaccaccccttagagccatgagagagacctgt 741
Db 6155 CGTAAGGAGTAGCCCGGGGAGTTTAAACCAACCCCTAGACCATGAGAAAGACCTGTC 6214
QY 742 gacctactgtcaagaagactgtatcctgtagccagtggttggcccgatgtctgaaagct 801
Db 6215 GCCTACCTGTCAAGAAAGCTGATGCTGTAGCCAGTGTGTCGCCATATGCTGTAAGAGCT 6274
QY 802 atcgcaactgtgcccatactggttcaagaagcgtgcaaaattgactttggagc-agaat 860
Db 6275 ATCCACACTGTGGCCATACGTGTCAAGAGCGCTGACAAATTTGATTTGGGACAAAGATAT 6334
QY 861 aacgtataagcccccatgattgtagaagacatgcttcggcagcccccagaacagatgat 920
Db 6335 AACGTATATAGCCCCCATGATGATGAGAACATGTTGCGGACGCCCAAGCCAGTGGAT 6394
QY 921 gaccaagcccgcaatgacccacttcaaaagcgttctctcaagagagaggttcaacttgc 980
Db 6395 GACCAAGCCCGCATGACCCCATGATCAAAAGCCTTCTCTACAGAGAGGGTCAAGCTTGGC 6454
QY 981 tcaacagcgcgtctcaacctgtcaactcttgcctggaagagagctgtaaacagtgac 1040
Db 6455 TCACACCAACCGCTCTCAACCCCTGCCACTTCTGCTGAAGAGAGCATGATCAACCGTGC 6514
QY 1041 tcatgattgcatcaactatgatttgaagagactgggttcgcgaagagacttcaagacat 1100
Db 6515 TCATGATTGCCATCACTATTGATTGATGAGAGACTGGGGTCCGCAAGGACCTTTACAGACAT 6574
QY 1101 accgtctagctgagaaatgtctaacctgttcaactgtacgtacgagacagctatg tgytgaag 1160
Db 6575 ACCGTGACTGAGAAAGTGCCTAACCTGTTCTGACGAGGAAAGCATGATGGTGGAGAG 6634
QY 1161 taagaagatggtctgggagcagctggttggagaggaaccgcagatctgggacagagct 1220
Db 6635 TAAAGAGATGGCTGGGGGGGGGTGGTGGAGGGGAGCCGACAGATCTGGGGCCAGCACCT 6694
QY 1221 gccggaaggaactcagcgcaaaagcgttgaagctcagtcctcagcgaagcttggagct 1280
Db 6695 GCCGGAGGAGACTTCAAGCAAAAAGGCTGAGCTCATGGCCCTCAAGGAGCTTTGGCGCT 6754
QY 1281 gccggaaggaatcacaataacatllacgagacaggtatagcttgcctttgcagctgaca 1340
Db 6755 GGCGGAAGGGAATTCATTAACATTTTATACGACAGCAGGATATCCTTTTGGAGATGCACA 6814
QY 1341 cgtacaaggggcatctatacaaaaggggtgtgcttaactcagcagagggagaaataa 1400
Db 6815 CGTACATGGGGCATCTATATAAACAAGGGGGTGTCTTACCTCAGCAGGAGGGGGAATATA 6874

QY 1401 gaacaaagaggaattcttaagcctatagaagcc 1434
Db 6875 GAACAAGAGAGAAATTCTTAAGCCTATTAGAACCC 6908

RESULT 5
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match 9.7%; Score 321; DB 3; Length 3482;
Best Local Similarity 99.2%; Pred. No. 1.7e-133;
Matches 521; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1767 cctgcagcagttggttaagaacatcccttatatgttctgaggtctaccaggggtgctga 1826
Db 6 cctgcagcagttggttaagaacatcccttatatgttctgaggtctaccaggggtgctga 65
QY 1827 ctcgtgtgtcaaacatctgtgtccctgcagctgtgtaattgtatccttccagatacc 1886
Db 66 ctgtgtgtgtcaaacatctgtgtccctgcagctgtgtaattgtatccttccagatacc 125
QY 1887 tccaagaaagagactaaggaggaagccaccagcgctcaactgaggaagtgaactca 1946
Db 126 tccaagaaagagactaaggaggaagccaccagcgctcaactgaggaagtgaactca 185
QY 1947 ggtaaagccggctaaataaggaacaaatatctatgttctgttagaacctttcagg 2006
Db 186 ggtaaagccggctaaataaggaacaaatatctatgttctgttagaacctttcagg 245
QY 2007 atggtagagagcttacttactaagaagaagactcaaccgtgtgtgtaagaataact 2066
Db 246 atggtagagagcttacttactaagaagaagactcaaccgtgtgtgtaagaataact 305
QY 2127 ttctgttgcaggttaagtcagggactggccaagatattggggattgatttgaacttga 2186
Db 366 ttctgttgcaggttaagtcagggactggccaagatattggggattgatttgaacttga 425
QY 2187 ttgtgcatagagccccaagctcagagacaggtagagagatgaatagaaacttaaa 2246
Db 426 ttgtgcatagagccccaagctcagagacaggtagagagatgaatagaaacttaaa 485
QY 2247 gaccttaccnaattgaccacagagactggcalttaattgagat 2291
Db 486 gaccttaccnaattgaccacagagactggcalttaattgagat 530

RESULT 6
US-09-376-781-7
; Sequence 7, Application US/09376781

Patent No. 6261806
GENERAL INFORMATION:
APPLICANT: Banerjee, Papia T.
APPLICANT: Pallence, Clive
APPLICANT: Andersson, Goran K.
TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
Patent No. 6261806
TITLE OF INVENTION: Use
FILE REFERENCE: 61750-267
CURRENT APPLICATION NUMBER: US/09/376,781
CURRENT FILING DATE: 1999-08-18
EARLIER APPLICATION NUMBER: 60/097,015
EARLIER FILING DATE: 1998-08-18
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 7
LENGTH: 445
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
OTHER INFORMATION: sequence of a portion of the PERV-A env gene
OTHER INFORMATION: corresponding to nucleotides 364-809 of Genbank
OTHER INFORMATION: No. 6261806 Y12238.
US-09-376-781-7

Query Match 8.0%; Score 265; DB 4; Length 445;
Best Local Similarity 99.5%; Pred. No. 6.5e-107;
Matches 435; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 2823 cccataaacctatctctcactcgttgtaactactactcgcgtacaggtataatla 2882
DB 1 cccataaacctatctctcactcgttgtaactactactcgcgtacaggtataatla 60
OY 2883 acagactcaaggaggagctccttgaggacctggtgacctgaattatgtcgccttc 2942
DB 61 acagactcaaggaggagctccttgaggacctggtgacctgaattatgtcgccttc 120
OY 2943 gatcagtaatccctggttcaatgacagcagcagcccccgatgtaactcgttctacg 3002
DB 121 gatcagtaatccctggttcaatgacagcagcagcccccgatgtaactcgttctacg 180
OY 3003 ggtttacgtttgcccagaccocccaaataatgaagaatatttggaatcctcagatt 3062
DB 181 ggtttacgtttgcccagaccocccaaataatgaagaatatttggaatcctcagatt 240
OY 3063 tcccttgaagaagaatgagctgcatatcttaattatg-agaattggaatgcccagtc 3121
DB 241 tcccttgaagaagaatgagctgcatatcttaattatg-agaattggaatgcccagtc 300
OY 3122 tctcagaagaagagtaagtactcttcttgaacaatcctacagttataatcaatt 3181
DB 301 tctcagaagaagagtaagtactcttcttgaacaatcctacagttataatcaatt 360
OY 3182 aattatggccatggagatggaagaatgacacagcggtgtaaaaaagatgtcgaat 3241
DB 361 aattatggccatggagatggaagaatgacacagcggtgtaaaaaagatgtcgaat 420
OY 3242 aagcaataaagctgtca 3258
DB 421 aagcaataaagctgtca 437

RESULT 7
US-09-075-272-1
Sequence 1, Application US/09075272
Patent No. 6136598

GENERAL INFORMATION:
APPLICANT: MILLER, A. DUSTY
APPLICANT: WOLGAMOT, GREG
APPLICANT: BONHAM, LYNN
TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL

TITLE OF INVENTION: PACKAGING CELL LINES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
City: San Francisco
STATE: California
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,272
FILING DATE: 08-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,140
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: POOR, Brian W.
REGISTRATION NUMBER: 32,928
REFERENCE/DOCKET NUMBER: 14538A-003710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 467-9600
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8655 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-075-272-1

Query Match 1.1%; Score 35; DB 3; Length 8655;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2203 caaagctcaagcaggtagaagatgtaagaaac 2237
DB 5249 CAAAGCTCAAGCAGGTAGAGATGAATAGAAC 5283
RESULT 8
US-09-352-990-7
Sequence 7, Application US/09352990
Patent No. 6255090
GENERAL INFORMATION:
APPLICANT: Famodu, Layo O.
APPLICANT: Orozco, Buddy
APPLICANT: Rafalski, Antoni
TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
FILE REFERENCE: BB-1191
CURRENT APPLICATION NUMBER: US/09/352,990
CURRENT FILING DATE: 1999-07-14
EARLIER APPLICATION NUMBER: 60/092,866
EARLIER FILING DATE: July 15, 1998
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Microsoft Office 97
SEQ ID NO 7
LENGTH: 836
TYPE: DNA
ORGANISM: Trillium aestivum
US-09-352-990-7

Query Match 0.9%; Score 30; DB 4; Length 836;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3291 tcactaaaaaaaaaaaaaaaaaaaaa 3320
|||||
Db 742 tcactaaaaaaaaaaaaaaaaaaaaa 771

RESULT 9
US-09-277-716-31
; Sequence 31, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(720)
; FEATURE:
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (Insert B)
US-09-277-716-31

Query Match 0.9%; Score 30; DB 4; Length 1279;
Best Local Similarity 100.0%; Pred. No. 0.00051;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3291 tcactaaaaaaaaaaaaaaaaaaaaa 3320
|||||
Db 1237 tcactaaaaaaaaaaaaaaaaaaaaa 1266

RESULT 10
US-08-798-000-3
; Sequence 3, Application US/08798000
; Patent No. 5766945
; GENERAL INFORMATION:
; APPLICANT: Miller, A. Dusty
; TITLE OF INVENTION: 10A1 RETROVIRAL PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,000
; FILING DATE: 12-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,564
; FILING DATE: 13-FEB-1996

; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14538A-002010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-798-000-3

Query Match 0.9%; Score 29; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2185 catlgtgcacagaccaccaagctcagg 2213
|||||
Db 96 catlgtgcacagaccaccaagctcagg 124

RESULT 11
US-09-248-335-27
; Sequence 27, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEF, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; CURRENT FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/924,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 27
; LENGTH: 1198
; TYPE: DNA
; ORGANISM: maize
US-09-248-335-27

Query Match 0.9%; Score 29; DB 3; Length 1198;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3292 cactaaaaaaaaaaaaaaaaaaaaa 3320
|||||
Db 1109 cactaaaaaaaaaaaaaaaaaaaaa 1137

RESULT 12
US-08-798-000-1
; Sequence 1, Application US/08798000
; Patent No. 5766945
; GENERAL INFORMATION:
; APPLICANT: Miller, A. Dusty
; TITLE OF INVENTION: 10A1 RETROVIRAL PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,000
FILING DATE: 12-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/011,564
FILING DATE: 13-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Patmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-002010
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-798-000-1

Query Match 0.9%; Score 29; DB 1; Length 1234;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2185 catctgcatagagaccaccaagctcag 2213
|||||
DB 96 CATGTGATACAGACCCCAAGCTCAG 124

RESULT 13
US-09-232-200-66
Sequence 66, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
TITLE OF INVENTION: FATY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21p3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 66
LENGTH: 2338
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-200-66

Query Match 0.9%; Score 29; DB 4; Length 2338;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3292 cactaataaaaaaaaaaaaaaaaaa 3320
|||||
DB 2295 cactaataaaaaaaaaaaaaaaaaa 2323

RESULT 14
US-09-232-197-66
Sequence 66, Application US/09232197A
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
TITLE OF INVENTION: FATY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21p3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 66
LENGTH: 2338
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-197-66

Query Match 0.9%; Score 29; DB 4; Length 2338;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3292 cactaataaaaaaaaaaaaaaaaaa 3320
|||||
DB 2295 cactaataaaaaaaaaaaaaaaaaa 2323

RESULT 15
US-09-011-745-3
Sequence 3, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Robin A
APPLICANT: Takeuchi, Yasuhiro
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
CURRENT FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GB96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
EARLIER FILING DATE: 1995-08-23
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 7308
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-3

Query Match 0.9%; Score 29; DB 4; Length 7308;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2185 cattgtgcatcacagaccaccaagctcag 2213
|||||
Db 6178 cattgtgcatcacagaccaccaagctcag 6206
|||||

Search completed: February 24, 2002, 07:23:41
Job time: 16030 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:23:41 ; Search time 197.25 Seconds

(without alignments)
9425.389 Million cell updates/sec

Title: US-09-171-553B-3
Perfect score: 8209
Sequence: 1 gtgtgtctacactgttgcc.....aaaaaaaaaaaaaaaa 8209

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTOUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2967	36.1	7333	4 US-08-766-528-2	Sequence 2, Appli
2	1466	17.9	3482	3 US-09-111-085-3	Sequence 3, Appli
3	704	8.6	8132	4 US-08-766-528-3	Sequence 3, Appli
4	457	5.6	2462	3 US-09-111-085-1	Sequence 1, Appli
5	375	4.6	8060	4 US-08-766-528-1	Sequence 1, Appli
6	278	3.4	278	4 US-09-376-781-10	Sequence 10, Appli
7	86	1.0	276	4 US-09-376-781-8	Sequence 8, Appli
8	35	0.4	8655	3 US-09-075-272-1	Sequence 1, Appli
9	29	0.4	735	1 US-08-798-000-3	Sequence 1, Appli
10	29	0.4	1234	1 US-08-798-000-1	Sequence 1, Appli
11	29	0.4	7308	4 US-09-011-745-3	Sequence 3, Appli
12	29	0.4	7308	4 US-09-011-745-4	Sequence 4, Appli
13	29	0.4	7616	4 US-09-011-745-2	Sequence 2, Appli
14	29	0.4	8302	1 US-08-258-420-13	Sequence 13, Appli
15	29	0.4	8332	3 US-08-850-961-1	Sequence 1, Appli
16	28	0.3	698	4 US-09-376-781-18	Sequence 18, Appli
17	28	0.3	949	5 PCT-US91-08254-1	Sequence 1, Appli
18	28	0.3	949	5 PCT-US91-08254-2	Sequence 2, Appli
19	28	0.3	1493	4 US-09-376-781-24	Sequence 24, Appli
20	28	0.3	1736	4 US-09-360-197-13	Sequence 13, Appli
21	28	0.3	2000	4 US-09-376-781-25	Sequence 25, Appli
22	28	0.3	2000	4 US-09-376-781-30	Sequence 30, Appli
23	27	0.3	1587	3 US-09-108-020-11	Sequence 11, Appli
24	26	0.3	843	2 US-08-831-570-2	Sequence 2, Appli
25	26	0.3	1098	3 US-09-248-335-35	Sequence 35, Appli
26	26	0.3	1275	3 US-08-725-532A-2	Sequence 2, Appli
27	26	0.3	4032	1 US-08-126-587C-8	Sequence 8, Appli

28	26	0.3	80595	4 US-09-078-294-3	Sequence 3, Appli
29	25	0.3	25	4 US-09-376-781-23	Sequence 23, Appli
30	25	0.3	945	4 US-09-268-364-3	Sequence 3, Appli
31	25	0.3	1000	4 US-09-357-251-19	Sequence 19, Appli
32	25	0.3	1001	1 US-08-728-259A-10	Sequence 10, Appli
33	25	0.3	1001	2 US-08-473-486-10	Sequence 10, Appli
34	25	0.3	1026	2 US-08-713-000-6	Sequence 6, Appli
35	25	0.3	1026	2 US-08-975-316-6	Sequence 6, Appli
36	25	0.3	1026	4 US-09-211-710-6	Sequence 6, Appli
37	25	0.3	1055	4 US-08-702-344-18	Sequence 18, Appli
38	25	0.3	1065	2 US-08-605-106-3	Sequence 3, Appli
39	25	0.3	1622	4 US-09-334-601-9	Sequence 9, Appli
40	25	0.3	1737	4 US-09-416-050A-7	Sequence 7, Appli
41	25	0.3	1737	4 US-09-664-800-7	Sequence 7, Appli
42	25	0.3	1737	4 US-09-665-309-7	Sequence 7, Appli
43	25	0.3	1737	4 US-09-661-569-7	Sequence 7, Appli
44	25	0.3	1826	2 US-08-808-931-15	Sequence 15, Appli
45	25	0.3	1826	3 US-08-808-931-15	Sequence 15, Appli

ALIGNMENTS

RESULT 1
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF INVENTION: AND METHODS OF USE
; CORRESPONDENCE ADDRESS: 74
; ADDRESS: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-2

Query Match 36.1%, Score 2967, DB 4, Length 7333;
Best Local Similarity 99.8%, Pred. No. 0;
Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 1 gtgtgtctacactgttgcccccagcgcttggaataaacctctgtgctgctc 60
|||||

Dh 11 GTGATGACAGCTGTGGGCCCCAGCGCTTGGAATAAATCCCTTGGTTTGATC 70
Qy 61 aagaccgtctcgtgagatgattggggtgcgcctctccagaccggagcagggat 120
Dh 71 AAGACCGCTCTTGTGATGATTTGGGTGGCTCTTCGACCCGGAGAGGGGAT 130
Qy 121 tgtcttctactgacctcattctgltgltggtccgggaaatcttcgacacccctta 180
Dh 131 TGTCTTACTGGGCTTTCATTTGGTGGCTGGCCGGAAATCTTCGACACCCCTTA 190
Qy 181 caccggagacagcacttggagtaaaagatccctcttggaaagltgtgtgtggtg 240
Dh 191 CACCCGAGAACCGACTTGAGGTAAAGGATCCCTTTGGAACATATGTGTGGCC 250
Qy 241 ggcgtctctgtgagatgtctgttctcgtgtgagcgcccttcggtcttcagcttct 300
Dh 251 GGGCTCTCTGTCTGAGGTCTGTGTGGGTGGATGCGGCTTTCGGTTTCAGTGTCT 310
Qy 301 ctccagaccgttaagagacttggagactgtgtatcagacagctgtctaggagagatca 360
Dh 311 CTCAGACCGTAAAGGACTGGAGAGACTGTGATCAGACAGTGTAAAGAGATCACAG 370
Qy 361 ccaaccctggggagacgcccgggaggttgggagagccagggacgctgtgtgtctact 420
Dh 371 CCACCCCTGGGGAGCGCCCGGAGGTGGGAGAGCCAGGAGCGCTGTGCTCCACT 430
Qy 421 gtctgtaagagagcaggtctgttctgttgaagcgaagctccctcccgcggtcgt 480
Dh 431 GTCGGTAAAGAGACCGAGTTCTGTGTGAAGCAAAAGCTTCCCTCCCGCCCTCCG 490
Qy 481 actcttctgctcgtctgttgaagacgagcgagctcggtgtgtctgtatctgtgtt 540
Dh 491 ACTCTTGGCTGCTGTGTGAAGCGCGAGCGGTGGCTGTGTGTGATCTTGTGT 550
Qy 541 ctgtctgtgtgtctgtctgtctgtgtgtctgtctgtctgtctgtctgtctgt 600
Dh 551 CTGTTCTGT 610
Qy 601 tgaactaccccccttgaattgaactcgaacatcgaactgaactgaactgaactga 660
Dh 611 TGACTACCCCCCTTGAATTGATCTGTGACCATTTGACGTGAATTAATCCAGGCT 670
Qy 661 atttctgaactcaggttaagaagagccttggcagactctgtgctctgtgaatg 720
Dh 671 ATTTGTCAGTTGAGTTAAAGAGGACTTGGCAGACTTCTGTGCTGTGAATGGCCA 730
Qy 721 catctgattgtgattgagcctcagagagggagccttaattcctgaattcgtgtg 780
Dh 731 CATTCGATGTTGGATGGCATCAGAGGGGACCTTTAATTCTGAATTAATCCGCTG 790
Qy 781 aggaacatcttctcagactgacccggtctcactcactcagagagccttatctcta 840
Dh 791 AGGCATCATTTTTCAGACTGGACCCGCTCTCATCTTATCTAGGAGCCCTTATCT 850
Qy 841 cgtgcaagatttggcagaagatcctcgcgcaltggttlaaacatctgtctaaata 900
Dh 851 CGTGGCAAGATTGGCAAGAGATCTCCGCATGGGTAAACATGGCTAAATTAACCA 910
Qy 901 gaaagcagagtcctcgaatcctgtgtctgtggagagaaacaaacactcgcgcaaa 960
Dh 911 GAAAGCAAGTCCCGAAATCTGGCTTGGAGAGAAAAAACAACACTGGCGGAAAA 970
Qy 961 tcgagccctctcctcgtactcaccagatcgaagagcgcgcgacttggcggaaacc 1020
Dh 971 TCGAGCCCTC-TCCTGTATCTACCCCGAGATCGAGAGCCCGCACTTGGCCGGA 1029
Qy 1021 caactgttcccccaccccttatacagacagaggtgtgtgtgtgtgtgtgtgt 1079
Dh 1030 CAACCTGTTCGCCACCCCTTATCCAGACAGGGGTGTGTGTGTGTGTGTGTGT 1089
Qy 1080 cctgaggtcctggt 1139
Dh 1090 CCTGGAGCTCGGT 1149
Qy 1140 ccggaagcagacagatcgcgatatccgcttgcacactatgcccctccatgcca 1199
Dh 1150 CCGAGCGGACAGACAGATCGGATTAATACCGCTCGCAGCCATGGCCCTCCATGCCA 1209
Qy 1200 gggggccaattgagccctccagatctgccccttctctcgcagatctcataatg 1259
Dh 1210 GGGGCGCAATGTGAGCCCTCCAGATATTTGGCCCTTTCTTGTGAGATCTTAT 1269
Qy 1260 aaactaacatccccccttctcagagatccccaacgctcagaggttggtaggtcc 1319
Dh 1270 AAACCTAACATCCCTTCTTCTGAGAGATCCCAACGCTTACGAGGTTGGTAGTCC 1329
Qy 1320 ctatgtctctcaccagctacttggagatgtgttcaacagctgtctgcagacactt 1379
Dh 1330 CTATGTCTCTACACGCTACTTGGATGATTTGCAACACGCTCTGAGACATCTTC 1389
Qy 1380 acaacgagagcagagagagatctgtttagaggttagaanaaatttccctgtgg 1439
Dh 1390 ACAACGAGAGAGAGAGAGATTTCTTGAAGGCTTAAAAAAAATTTCTGGGCG 1449
Qy 1440 gggcagccacgcaattgcaaaatagatltgacatgggatttcccttaccgcgcggt 1499
Dh 1450 GGGCAGCCACGCAATGTGCAAAATGAGATTGACATGGATTTCTTACTGCCCCGT 1509
Qy 1500 tgggactaacacagcgttgaaggttagggagcttgaatctatcgcaggtctgtg 1559
Dh 1510 TGGGACTAACACAGCGCTGAAGGTAGGAGAGACTTGAATAATCTATCGCCAG 1569
Qy 1560 ggggtctcggggcgctcgaagagcgccactaaattgcttaaggttagagaggtg 1619
Dh 1570 GGGGTCTCGGGGGGCTCAAGAGCGCCCTAATTTGGCTTAAAGTAAAGAGGTATG 1629
Qy 1620 cagggacggaagaaactccctcgttattcttgaagagctcaltgaagcctcagc 1679
Dh 1630 CAGGACCGAAGCACTCCCTCGGTATTCTGTAGAGGCTCATGGAAGCCTTCA 1689
Qy 1680 ttcaaccccttctgactcactcagagcccaagaagcctcagltgcccctgct 1739
Dh 1690 TTCACCCCTTTTGTGATCTTACCTCAGAGGCCCAAGAAAGCTCAGTGGCCCT 1749
Qy 1740 gggcagctcgtctgatatcagagaaagaaatcagagacgtggaaggtttaca 1799
Dh 1750 GGGCGAGTGGCTCTTGATATCAGAGAAACCTTCAAGACCTGGAAGGGTTAC 1809
Qy 1800 gattacgttgaactagtagagagagcagagaggtgttlaacagaaagtagacag 1859
Dh 1810 GAGTTACGTGATCTGTAGAGAGGACAGAAAGGTATTTACAGAGGACACAGAG 1869
Qy 1860 gagaag 1919
Dh 1870 GAGAGAGAACAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1929
Qy 1920 caaggaagaaatttgaactaagatttggcgcgaatttgaaggaagcagcagagag 1979
Dh 1930 CAAGAGAAATTTTGAATTAATTTGAGGAGGCTTGAAGGAGAGACAGAGAG 1989
Qy 1980 agagagagagattttagaanaattaggtlcaaggccctagacagltcaaggaaact 2039
Dh 1990 AGAGAGAGATTTTGAATAATTTAGGTACAGGCTTGAAGAGAGAGAGAGAGAG 2049
Qy 2040 agagaccacactcgaagagacaggtgtgtgtatgttgaagaaagagacactggc 2099
Dh 2050 AGGACCCCACTCGAAGAGAGCAATGTGTGCTTAAAGAAAAAGACACTGGCGA 2109
Qy 2100 aactgcccagaag 2159
Dh 2110 AACTGCCCAAG 2169
Qy 2160 taggagagagaggttctgagacccctcccgagcccgaggttaacttgaaggttga 2219
Dh 2170 TAGGGAGAGAGGAGGTTTGGAGACCCCTCCGAGGCCAGGGTAACTTTGAAGGT 2229

Dh 1350 cgatctctccagagccgggttaaatctctcttctgaaltccgcccgggcaagatca 1409
QY 6119 aatctgataactatataagaatgaagctgctcccatcagacttagatataaga 6178
Dh 1410 aagctgataactatataagaatgaagctgctcccatcagacttagatataaga 1469
QY 6179 taagttcaactg--aaagaaacagaaataatcacaagtgataatglatgact 6236
Dh 1470 taagttcaactgaaagaaacagaaataatcacaagtgataatglatgact 1529
QY 6237 ggggaatgctttttat-tatatggcggggaaggggtccacttaacattccgctta 6295
Dh 1530 ggggaatgctttttatataatgacgggggaaggggtccacttaacattccgctta 1589
QY 6296 ggaataagagcgggagacagacccctctggtcaatggtgacccgataaagactgctgaac 6355
Dh 1590 ggaataagagcgggagacagacccctctggtcaatggtgacccgataaagactgctgaac 1649
QY 6356 aggggcccccgccctggaacccacagataactgctggtgccccataactccgctgac 6415
Dh 1650 aggggcccccgccctggaacccacagataactgctggtgccccataactccgctgac 1709
QY 6416 ggcctgacataacacagccgctcagacagatcacactggtatcttccatacaacagc 6475
Dh 1710 ggcctgacataacacagccgctcagacagatcacactggtatcttccatacaacagc 1769
QY 6476 ctagaactcccaaggctgttctctgttaagaacaggaagagactcttcagctccacag 6535
Dh 1770 ctagaactcccaaggctgttctctgttaagaacaggaagagactcttcagctccacag 1829
QY 6536 gaacttccaaagccatcaactccacagaccccgatgacactctctctgttggctgttc 6555
Dh 1830 gaacttccaaagccatcaactccacagaccccgatgacactctctctgttggctgttc 1889
QY 6596 talccctcagggccctcttataatgagggatggtcctaaagaaatcactgacga 6655
Dh 1890 talccctcagggccctcttataatgagggatggtcctaaagaaatcactgacga 1949
QY 6656 aagaagcagaaatcaatgataatgagggatggtcctaaagaaatcactgacga 6715
Dh 1950 aagaagcagaaatcaatgataatgagggatggtcctaaagaaatcactgacga 2009
QY 6716 ccgggaaggggagcagatcagtaagaaagctcccccatacccaacacacttgcatagta 6775
Dh 2010 ccgggaaggggagcagatcagtaagaaagctcccccatacccaacacacttgcatagta 2069
QY 6776 ctgtgttctatgagcagggcctcagaaatcagtaatgagggatggtcctaaagaaatc 6835
Dh 2070 ctgtgttctatgagcagggcctcagaaatcagtaatgagggatggtcctaaagaaatc 2129
QY 6836 gggcagatcaatcaggggttaacccctgtgttccacactcagcttccacacacacacaa 6895
Dh 2130 gggcagatcaatcaggggttaacccctgtgttccacactcagcttccacacacacacaa 2189
QY 6896 attctgtgctatgcttccaaatcgttcccccagtgctactacactcctgaggaagtgctcc 6955
Dh 2190 attctgtgctatgcttccaaatcgttcccccagtgctactacactcctgaggaagtgctcc 2249
QY 6956 ttgataatgataatcagttacacagacaaagaaagaaacccgatacccttaacactag 7015
Dh 2250 ttgataatgataatcagttacacagacaaagaaagaaacccgatacccttaacactag 2309
QY 7016 ctgtaatgctcggattaggaagcggcgttgcgttaagaaacaggaagctgcccgtgaca 7075
Dh 2310 ctgtaatgctcggattaggaagcggcgttgcgttaagaaacaggaagctgcccgtgaca 2369
QY 7076 caagacacacagcagctcaggaagaaagagctgtgtgactacacagcagctgacagaa 7135
Dh 2370 caagacacacagcagctcaggaagaaagagctgtgtgactacacagcagctgacagaa 2429
QY 7136 tccgaagccttaagagagctgtttagcaacactgaagagctccctgactcttgcctgaag 7195
Dh 2436 tccgaagccttaagagagctgtttagcaacactgaagagctccctgactcttgcctgaag 2489

QY 7196 tggcttctacagaacccggaggatagactcgtcttctaaagaaaggtggtatggt 7255
Dh 2490 tggcttctacagaacccggaggatagactcgtcttctaaagaaaggtggtatggt 2549
QY 7256 cagccttaaaagaaatggttgcctctctctctctctctctctctctctctctctctct 7315
Dh 2550 cagccttaaaagaaatggttgcctctctctctctctctctctctctctctctctctct 2609
QY 7316 tgaacaaagccttaagaaanaatctaaagagcgttgaaggggaagagagcgtgacaggg 7375
Dh 2610 tgaacaaagccttaagaaanaatctaaagagcgttgaaggggaagagagcgtgacaggg 2669
QY 7376 ggttgaagagatggttcaacaggtcctctctctctctctctctctctctctctctct 7435
Dh 2670 ggttgaagagatggttcaacaggtcctctctctctctctctctctctctctctctct 2729
QY 7436 ggcctcagtagtct 7495
Dh 2730 ggcctcagtagtct 2789
QY 7496 ttgccttcttgaagagagatgagtgagtcagtcagtcagtcagtcagtcagtcagtcag 7555
Dh 2790 ttgccttcttgaagagagatgagtgagtcagtcagtcagtcagtcagtcagtcagtcag 2849
QY 7556 aagggccttctgagccaggaagaaactgaccccttaagcctccagttcctaagatgaact 7615
Dh 2850 aagggccttctgagccaggaagaaactgaccccttaagcctccagttcctaagatgaact 2909
QY 7616 ataaacagaacagaagatggtggaatgaaagatgaaatgaaatgaaatgaaatgaaat 7675
Dh 2910 ataaacagaacagaagatggtggaatgaaagatgaaatgaaatgaaatgaaatgaaat 2969
QY 7676 ccaggaggttaataaagcctctaaatgccccgaatctaacacactgctgctgacgt 7735
Dh 2970 ccaggaggttaataaagcctctaaatgccccgaatctaacacactgctgctgacgt 3029
QY 7736 aaatggtgaagagcagacacttccatggttccagggcgtgctactcctgctcctaagta 7795
Dh 3030 aaatggtgaagagcagacacttccatggttccagggcgtgctactcctgctcctaagta 3089
QY 7796 ataaacagaacagaatggttgaatcagtcagtcagtcagtcagtcagtcagtcagtcag 7855
Dh 3090 ataaacagaacagaatggttgaatcagtcagtcagtcagtcagtcagtcagtcagtcag 3149
QY 7856 aagaagaatggttgaatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 7915
Dh 3150 aagaagaatggttgaatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 3209
QY 7916 tgcacagagagccagcagatgagcagcctcagagcctattttaaataatggttgcacag 7975
Dh 3210 tgcacagagagccagcagatgagcagcctcagagcctattttaaataatggttgcacag 3269
QY 7976 agcgcggcctcctcagatctttaaataatggttgcacagcctcagagcctattttaaata 8034
Dh 3270 agcgcggcctcctcagatctttaaataatggttgcacagcctcagagcctattttaaata 8094
QY 8035 taaatattggttgaagcagcagccttgcgttgcacacccataaaagcgttccagat 8094
Dh 3330 taaatattggttgaagcagcagccttgcgttgcacacccataaaagcgttccagat 3389
QY 8095 tccgcactcggggcggcagtcctctacacccctcgttgcgttgcacacccataaaagcgt 8154
Dh 3390 tccgcactcggggcggcagtcctctacacccctcgttgcgttgcacacccataaaagcgt 3449
QY 8155 gcttgaataataaactcct 8187
Dh 3450 gcttgaataataaactcct 3482

RESULT 3
US-08-766-528-3
; Sequence 3, Application US/08766528

Db 4470 GACCGGGTGGCAAGCGAGCCAGGCTTAACTTGGCTTAAATGAATGCC 4529
QY 4533 aaagcccaagaccagacaglacacccatagaagacatggcaagataaaagata 4592
Db 4530 AAAGCCCGAGAACCCAGACGACATACACCTACAAACATGCGCAAGATAAAAAGATA 4589
QY 4593 gaccatgtctctgaagaccgagggagactgtataccatataatggaaagaatctg 4652
Db 4590 GACCAAGTCTCTGAGACTCCGGAAGGAGCTGCTATACCTAGATGGGAAGAAATCTTG 4649
QY 4653 ccccaacaagaagggttagaatatgtccacaagaatataatgttaaccacccatagaact 4712
Db 4650 CCCACAAAGAGGGTTAGATATGTCAACAGATACATGTCTAACCCACCTAGGAAGT 4709
QY 4713 aaacactgacgaagtgtgacagacacccatataatgtctgaagctaccagagtg 4772
Db 4710 AAACACCTGCAGCAGTGTGACAGAACATCCCTTATCATCTTGTGAGGCTACAGAGAGTG 4769
QY 4773 gctgactcagtggtcaaacattgtgacctgacagctgttaatgttaactccacaga 4832
Db 4770 GCTGACTCGGTGTCAAAACATTTGTGCTCCCTGCAGCTGTTATATGCTTATCTCCAGA 4829
QY 4833 ataccctcagaaagaaagacaaagggaagccacccagcgctactgggaagtgagactc 4892
Db 4830 ATGCTCCAGGGAAGACACTAAGGGAAGCCACCCAGCCCTCACTGGGAGTGGACTTC 4889
QY 4893 actgaggtlaagcccgctaaataacgaaacaaatatactatgtgttttgaagacatt 4952
Db 4890 ACTGAGGTAAAGCCGGCTTAATACGGAACCAATACCTATTGTGTTTGTGAGACACTTT 4949
QY 4953 tcaagatgggttagagcttaactactaagaagaagactcaacggtgtgtgaagaa 5012
Db 4950 TCAGGATGGTAAAGGCTTACTACTAAGAAAGACACTTCAACCTGTGTGCTTAAAAAA 5009
QY 5013 atactggaagaaattttccaaagatttgaaataccaaagttaaggtcagacaaatgt 5072
Db 5010 ATACTGGAAGAAATTTTCCAAATTTTGAATACCTAAGGTANTAGGCTCAGACAAATGGT 5069
QY 5073 ccaagcttcgtgccaggttaagtcaggagactggccaagataatgggaattgaa 5132
Db 5070 CCAGCTTTTGTGTCGCCAGTAAAGTCAAGGAGCTGGCCCAAGATATTGGGGATTGTTGAAA 5129
QY 5133 ctgcatgtgcatagacagaccacaagctcaagacagatgaagaagataagaacatt 5192
Db 5130 CTGCATTGTGCTACAGACCCCAAGCTCAGAGAGTAGAGAGATGAATGAACCAATT 5189
QY 5193 aaagagacccttac 5206
Db 5190 AAAGAGACCCCTTAC 5203

RESULT 4
US-09-111-085-1
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111.085
; EARLIER APPLICATION NUMBER: 1998-07-07
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2462
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-1

Query Match 5.6%; Score 457; DB 3; Length 2462;
Best Local Similarity 99.8%; Pred. No. 4.1e-187;
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7335 gttagaagagctcgaaggaagaaagagctaccaggggtgtgttgaagaaagttca 7394
Db 1938 gttagaagagctcgaaggaagaaagagctaccaggggtgtgttgaagaaagttca 1997
QY 7395 caggtctcttgatgaacaccccttctgtccttgacggggcccttagtactgtct 7454
Db 1998 caggtctcttgatgaacaccccttctgtccttgacggggcccttagtactgtct 2057
QY 7455 cctgtacttaacagttgggcttgcttaataagttgtgtgttgaagaaag 7514
Db 2058 cctgtacttaacagttgggcttgcttaataagttgtgtgttgaagaaag 2117
QY 7515 agtgaagtcagtcagatattgtacttaggcaacagtaacaaagccttcgaagcaag 7574
Db 2118 agtgaagtcagtcagatattgtacttaggcaacagtaacaaagccttcgaagcaag 2177
QY 7575 agaactgaaccttagccttccagttctaaattagaactataacaaagaagaagtg 7634
Db 2178 agaactgaaccttagccttccagttctaaattagaactataacaaagaagaagtg 2237
QY 7635 gggatgaagatgaataatgaaccttaacctccagaaacccaggaagttataaag 7694
Db 2238 gggatgaagatgaataatgaaccttaacctccagaaacccaggaagttataaag 2297
QY 7695 ctctaattgccccgaattacagacccctgtcgtcgtcgaagtaaatagttgaagtcaca 7754
Db 2298 ctctaattgccccgaattacagacccctgtcgtcgtcgaagtaaatagttgaagtcaca 2357
QY 7755 ctctaattgccccgaattacagacccctgtcgtcgtcgaagtaaatagttgaagtcaca 7814
Db 2358 ctctaattgccccgaattacagacccctgtcgtcgtcgaagtaaatagttgaagtcaca 2417
QY 7815 cttaactgtactcgtgactcgttaaac 7842
Db 2418 cttaactgtactcgtgactcgttaaac 2445

RESULT 5
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965

REFERENCE/DOCKET NUMBER: MCP-038CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)227-5941
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8060 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-766-528-1

Query Match 4.6%; Score 375; DB 4; Length 8060;
 Best Local Similarity 98.7%; Pred. No. 6,7e-152;
 Matches 1395; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 2973 gatgcctcttcgctgagattacaacccactagccaacacttttgctcgaatg 3032
 |||||||
 Db 5495 GATGCCTTCTTCGCTGAGATTACACCCACTAGCCAACTTTTGCTTCGATG 5554
 QY 3033 agagatccaggtacgggaagaaacgggagctcaactggaaccgactgcccgaaggttc 3092
 |||||||
 Db 5555 AGAGATCCAGGTACGGGAAGAACCGGGAGCTCACCCTGGACCCGACTGCCCAAGGGTTC 5614
 QY 3093 aaaaacccccggacatctttgagaaagccctacacagagactggccaacttcaagatc 3152
 |||||||
 Db 5615 AAGAACTCCCGACCATCTTGTGACGAAGCCCTACACAGAACCTGGCCAACTTCAGAGATC 5674
 QY 3153 caaacaccctcaggtagaccctccctcagtagatgacgtcttcgacggagagccac 3212
 |||||||
 Db 5675 CAACACCTCAGGTGAGACCCCTCCACGATGAGTACCTGCTTGGGGGAGGACAC 5734
 QY 3213 aaacagagatgcttagaaggtacgaaggaactacgtctgaattctcgaactaggtac 3272
 |||||||
 Db 5735 AAACAGGACTGCTTAGAAGGACGAAGGACCTACTGCTGAATTTGCTGACCTAGGCTAC 5794
 QY 3273 agagcccttgcttaagaaagccagatttgcagagagaagtaacacttggggtacat 3332
 |||||||
 Db 5795 AGAGCCCTCTCTAAGAAAGGCCAGATTGTGAGAGAGAGGTAAACATACTTGGGGTACAGT 5854
 QY 3333 ttgagggcgagcgatgctgtagcaggaagcaggaagaaactgagtcacgaataccg 3392
 |||||||
 Db 5855 TTAGGGAGGCGGACGATGCTGAGAGAGGACGAGAAACTGTAGTCGATACCG 5914
 QY 3393 gccccaacacagccaacaaagtgaagagtttttgggagcagctggaatttgcagactg 3452
 |||||||
 Db 5915 GCCCAACACACAGCCAAACAAATGAGAGATTTTGGGGACAGCTGATTTTGCACAGCTG 5974
 QY 3453 tggatcccgaggttttgagacttagagcccaactctacccgttaacccaagaagaagg 3512
 |||||||
 Db 5975 TGGATCCCGGGGTTTGGAGCTTAGAGCCCACTCTACCCGCTAACCAAGAAAGAGG 6034
 QY 3513 ggaattccggggctccttgagcaccagaagcaattgagtcttcaaaaagccctgctg 3572
 |||||||
 Db 6035 GAATTTCCCTGGGCTCTGTGACACGAAGAGCATTTGATGCTATCAAAAAGGCCGCTG 6094
 QY 3573 agcgacactgctctgacctccctgagcttaactaaaccttacccttlatgtgagtag 3632
 |||||||
 Db 6095 AGCGACACTCTCTGTGGCCCTCCCTGAGCTAATTAACCCCTTTACCTTATGTGGATGAG 6154
 QY 3633 cgttaaggaatagcccgaggaagtttaacccaacaccttgagacatggagaggaactgtt 3692
 |||||||
 Db 6155 CGTAGGGAGTAGCCCGGAGATTTTAACCNAACCTTAGACCATGAGAAAGACCTGTGC 6214
 QY 3693 gctcactgtcaagaagacttgaatcctgtagcagagtggttggccgttatgtctgaaggt 3752
 |||||||
 Db 6215 GCTCACTGTCAAGAAGAGCTGATCTGTAGCCAGTGGTGGCCCATATGCTCGAAGGCT 6274
 QY 3753 atcgacagtgtgacatactgtgtcaaggaagcgtgacaacttgacttggagc-agaatat 3811
 |||||||
 Db 6275 ATCGACAGTGTGACATAGTGTCAAGAGCGCTGACAAATTTGACTTTGGGACAAAGATAT 6334

QY 3812 aactgtaatagcccccatgcatltggagaacatcgttcggcagccccccagacagatgat 3871
 |||||||
 Db 6335 AACTGTAATAGCCCCCATGATTTGGAAACATCTTTGGGACAGCCCCAGACCATGAT 6394
 QY 3872 gaccacgcccgcgatgacccaactlacaagcgtgtcttcacagaagagtgtaacttcgc 3931
 |||||||
 Db 6395 GACCAAGCGCCGATGACCCACTATCAAAAGCGTGTCTTCACAGAGAGGTCACGTGCG 6454
 QY 3932 tccaccagcgcctcgaacctgtccactctctcgtcctgaaagagactgtgaaccagtac 3991
 |||||||
 Db 6455 TCCACCAACCGCTTCAACCCCTGCCACCTTGTGCTTAAAGACTGATGAAACAGTAC 6514
 QY 3992 tcatattgccaatcaacttgaattgaagaaactgggttcgcaagagcccttaagacat 4051
 |||||||
 Db 6515 TCATATTGGCATCACTATTGATTTGAGAGACTGGGTCCGACAGACCTTACAGACAT 6574
 QY 4052 accgctgactggaagatgctaaacctgtgtcaactgacggaagcagctatgtgttgaag 4111
 |||||||
 Db 6575 ACCGCTGACTGAGAAAGTGTAACTGGTTCACCTGACGGAAGCACCTATGTGTGAAG 6634
 QY 4112 taagagatgcttggggcgaggtgtgaaagggaacccgacagatctggccaagcagct 4171
 |||||||
 Db 6635 TAAAGAGATGGCTGGGGCGCGGTGTGACGCGACCGACGATCTGGCCAGACGCT 6694
 QY 4172 gccggaagaaactcagcgaagagcgtgagcttaaggccctcaacgcaagcttgcgct 4231
 |||||||
 Db 6695 GCCGGAGGAACTTACGACAAAGAGCTGAGCTATGGCCCTACAGCAAGCTTGGGCT 6754
 QY 4232 ggcggaaggaatccataaactlatacgcagacagagatgcttgcgactgcaca 4291
 |||||||
 Db 6755 GSCCGAAGGAATCCATTAAACATTTATACGAGACAGAGTATGCTTTGGACATGCACA 6814
 QY 4292 cgtacaaggggcatctataaacaagggtgtgttccttcagaaggaaggaataaa 4351
 |||||||
 Db 6815 CGTACATGGGCGCATCTATAAACAAGGGGTTGCTTACCTCAGCAGGAGGGAATAAA 6874
 QY 4352 gaacaaagagaatcttaagcctaatagaacc 4385
 |||||||
 Db 6875 GAACAAGAGGAATTTTAAGCTTATAGAACCC 6908

RESULT 6
 US-09-376-781-10
 ; Sequence 10, Application US/09376781
 ; Patent No. 6261806
 ; GENERAL INFORMATION:
 ; APPLICANT: Banerjee, Papia T.
 ; APPLICANT: Patience, Clive
 ; APPLICANT: Andersson, Goran K.
 ; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
 ; Patent No. 6261806
 ; FILE REFERENCE: 61750-267
 ; CURRENT APPLICATION NUMBER: US/09/376,781
 ; CURRENT FILING DATE: 1999-08-18
 ; EARLIER APPLICATION NUMBER: 60/097,015
 ; EARLIER FILING DATE: 1998-08-18
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Patentl Ver. 2.0
 ; SEQ ID NO 10
 ; LENGTH: 278
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
 ; OTHER INFORMATION: sequence of part of poEV env gene corresponding to
 ; OTHER INFORMATION: residues 6057-6333 of WO 97/40167 for comparison.
 US-09-376-781-10

Query Match 3.4%; Score 278; DB 4; Length 278;
 Best Local Similarity 100.0%; Pred. No. 4.1e-110;

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/798,000
;; FILING DATE: 12-FEB-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/011,564
;; FILING DATE: 13-FEB-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Parmelee, Steven W.
;; REGISTRATION NUMBER: 31,990
;; REFERENCE/DOCKET NUMBER: 14538A-002010
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 206-467-9600
;; TELEFAX: 415-576-0300
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 735 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-08-798-000-3

Query Match 0.4%; Score 29; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5136 catgtgcatacagaccaccaagctcag 5164
|||
Db 96 CATTGTGCATACAGACCCCAAGCTCAGG 124

RESULT 10
US-08-798-000-1
; Sequence 1, Application US/08798000
; Patent No. 5766945
; GENERAL INFORMATION:
; APPLICANT: Miller, A. Dusty
; TITLE OF INVENTION: 10A1 RETROVIRAL PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,000
; FILING DATE: 12-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,564
; FILING DATE: 13-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14538A-002010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear

; MOLECULE TYPE: cDNA
US-08-798-000-1

Query Match 0.4%; Score 29; DB 1; Length 1234;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5136 catgtgcatacagaccaccaagctcag 5164
|||
Db 96 CATTGTGCATACAGACCCCAAGCTCAGG 124

RESULT 11
US-09-011-745-3
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-3

Query Match 0.4%; Score 29; DB 4; Length 7308;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5136 catgtgcatacagaccaccaagctcag 5164
|||
Db 6178 catgtgcatacagaccaccaagctcag 6206

RESULT 12
US-09-011-745-4
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 7308
; TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
OTHER INFORMATION: construct
US-09-011-745-4

Query Match 0.4%; Score 29; DB 4; Length 7308;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5136 catgtgcatacagaccccaagctcagg 5164
|||||
DB 6178 catgtgcatacagaccccaagctcagg 6206

RESULT 13
US-09-011-745-2
Sequence 2, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Robin A
APPLICANT: Takeuchi, Yasuhiro
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
CURRENT FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GH96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2
LENGTH: 7616
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
OTHER INFORMATION: construct
US-09-011-745-2

Query Match 0.4%; Score 29; DB 4; Length 7616;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5136 catgtgcatacagaccccaagctcagg 5164
|||||
DB 5339 catgtgcatacagaccccaagctcagg 5367

RESULT 14
US-08-258-420-13/C
Sequence 13, Application US/08258420
Patent No. 5710037
GENERAL INFORMATION:
APPLICANT: Nienhuis, Arthur W.
APPLICANT: Vanin, Elio F.
TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Vectors
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,420
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Olstein, Elliot M.
REGISTRATION NUMBER: 24,025
REFERENCE/DOCKET NUMBER: 271010-208
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 8202 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: viral genome
US-08-258-420-13

Query Match 0.4%; Score 29; DB 1; Length 8202;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5136 catgtgcatacagaccccaagctcagg 5164
|||||
DB 2535 catgtgcatacagaccccaagctcagg 2507

RESULT 15
US-08-850-961-1
Sequence 1, Application US/08850961
Patent No. 6013517
GENERAL INFORMATION:
APPLICANT: Respass, James G.
APPLICANT: De Polo, Nicholas J.
APPLICANT: Chada, Sunil
APPLICANT: Sauter, Sybille
APPLICANT: Bodner, Mordechai
APPLICANT: Driver, David A.
TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation, Intellectual Property - R440
STREET: P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/850,961
FILING DATE: 05-MAY-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kruse, No. 6013517/man J.
REGISTRATION NUMBER: 35,235
REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3520
TELEFAX: (510) 655-3542

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-850-961-1

```

```

Query Match          0.4%; Score 29; DB 3; Length 8332;
Best local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5136 catlgtgcatcacagaccacaagctcag 5164
    ||||||||||||||||||||||||||||
DB 5232 CATTTGTGCAATACAGACCCCAAGCTCAG 5260

```

Search completed: February 24, 2002, 07:30:07
 Job time: 16416 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on:      February 24, 2002, 07:30:07 ; Search time 197.25 Seconds
              (without alignments)
              3394.012 Million cell updates/sec
```

```
Title: US-09-171-553B-9
Perfect score: 2956
Sequence: 1 tgccttttaggttagaac.....aaaaaaaaaaaaaaaaaa 2956
```

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

word size

Total number of hits satisfying chosen parameters: 702406

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

post-processing: Listing first 45 summaries

Database :

```
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfile1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	length	DB	ID	Description
1	1743	59.0	3482	3	US-09-111-085-3	Sequence 3, Appli
2	813	27.5	7333	4	US-08-766-528-2	Sequence 2, Appli
3	457	15.5	2462	3	US-09-111-085-1	Sequence 1, Appli
4	147	5.0	276	4	US-09-376-781-8	Sequence 8, Appli
5	137	4.6	278	4	US-09-376-781-10	Sequence 10, Appli
6	98	3.3	8132	4	US-08-766-528-3	Sequence 3, Appli
7	91	3.1	8060	4	US-08-766-528-1	Sequence 1, Appli
8	28	0.9	1493	4	US-09-376-781-24	Sequence 24, Appli
9	28	0.9	1736	4	US-09-360-197-13	Sequence 13, Appli
10	28	0.9	2000	4	US-09-376-781-25	Sequence 25, Appli
11	28	0.9	2000	4	US-09-376-781-30	Sequence 30, Appli
12	27	0.9	1587	3	US-09-108-020-11	Sequence 11, Appli
13	26	0.9	843	2	US-08-831-570-2	Sequence 2, Appli
14	26	0.9	1098	3	US-09-248-335-35	Sequence 35, Appli
15	26	0.9	1275	3	US-08-725-532A-2	Sequence 2, Appli
16	26	0.9	4032	1	US-08-126-587C-8	Sequence 8, Appli
17	26	0.9	80595	4	US-09-078-294-3	Sequence 3, Appli
18	25	0.8	25	4	US-09-376-781-33	Sequence 23, Appli
19	25	0.8	698	4	US-09-376-781-18	Sequence 18, Appli
20	25	0.8	945	4	US-09-268-364-3	Sequence 3, Appli
21	25	0.8	1000	4	US-09-357-251-19	Sequence 19, Appli
22	25	0.8	1001	1	US-08-728-259A-10	Sequence 10, Appli
23	25	0.8	1001	2	US-08-473-486-10	Sequence 10, Appli
24	25	0.8	1026	2	US-08-713-000-6	Sequence 6, Appli
25	25	0.8	1026	2	US-08-975-316-6	Sequence 6, Appli
26	25	0.8	1026	4	US-09-211-710-6	Sequence 6, Appli
27	25	0.8	1055	1	US-08-702-344-18	Sequence 18, Appli

ALIGNMENTS

28	25	0	8	1062	2	US-08-605-106-3	Sequence 9, Appl
29	25	0	8	1622	4	US-09-334-601-9	Sequence 9, Appl
30	25	0	8	1737	4	US-09-416-050A-7	Sequence 7, Appl
31	25	0	8	1737	4	US-09-664-800-7	Sequence 7, Appl
32	25	0	8	1737	4	US-09-665-309-7	Sequence 7, Appl
33	25	0	8	1737	4	US-09-661-563-7	Sequence 15, Appl
34	25	0	8	1826	2	US-08-808-931-15	Sequence 15, Appl
35	25	0	8	1826	3	US-08-808-332-15	Sequence 15, Appl
36	25	0	8	1826	3	US-09-050-603A-15	Sequence 15, Appl
37	25	0	8	1826	3	US-09-102-420B-15	Sequence 15, Appl
38	25	0	8	1889	3	US-08-861-747-3	Sequence 3, Appl
39	25	0	8	1957	4	US-09-352-990-11	Sequence 11, Appl
40	25	0	8	2158	1	US-07-602-608-1	Sequence 1, Appl
41	25	0	8	2158	1	US-08-261-578-1	Sequence 1, Appl
42	25	0	8	2351	1	US-08-168-091A-1	Sequence 1, Appl
43	25	0	8	3212	4	US-08-697-954-1	Sequence 1, Appl
44	25	0	8	4337	3	US-09-187-040-1	Sequence 1, Appl
45	24	0	8	40	4	US-09-306-290-30	Sequence 30, Appl

```

RESULT 1
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
; US-09-111-085-3

```

	Query Match	59.0%;	Score 1743;	DB 3;	Length 3482;
	Best Local Similarity	99.3%	Pred. No. 0;		
	Matches 2913;	Conservative	0;	Indels 21;	Gaps 1;
QY	1 tgccttttagaggtttagaacaccccttgacagatttggtgctgaaccctatgatgtcct	60			
Db	548 tgccttttaagggttagaacaccccttgacagatttggtgctgaaccctatgatgtcct	607			
QY	61 acggggagaccccccccgttgtgabaatatgcttcctgtacataagtctgatgtgcgtctt	120			
Db	608 acggggagaccccccccgttgtgabaatatgcttcctgtacataagtctgatgtgcgtctt	667			
QY	121 cccaacctttgttctcttagctcaagaagcgcttcgaattggttgagacaagagcttgaagc	180			
Db	668 cccaacctcttctctcttagctcaagaagcgcttcgaattggttgagacaagagagcttgaagc	727			
QY	181 agctccggagagagctactaaagagagagagacttgcgaagttccaatcgtcttccaaattg	240			
Db	728 agctccggagagagctactaaagagagagagacttgcgaagttccaatcgtcttccaaattg	787			
QY	241 gagattcacgtcatgtttaagcgcacacgtgcacggaaaaccttcgaagctcgtgttgaaggagc	300			
Db	788 gagattcacgtcatgtttaagcgcacacacgtgcacggaaaaccttcgaagctcgtgttgaaggagc	847			
QY	301 cttaattcgttaattttgaccacaccaaagcgcttgaagaatcgaagaatctccaaacttga	360			
Db	848 cttaattcgttaattttgaccacaccaaagcgcttgaagaatcgaagaatctccaaacttga	907			

QY	361	tcatgatcccaacgftaaagcgtgagccaaactcccgatctgggttgagagccggaaga	420
Db	908	tccatgtatcccaacgftaaagcgtgagccaaactcccgatctgggttgagagccggaaga	967
QY	421	ctgagaatcccaactaaagcttcgcctcccaactgcgcgtgttcccttaactctcaaaataatccc	480
Db	968	ctgagaatcccaactaaagcttcgcctcccaactgcgcgtgttcccttaactctcaaaataatccc	1027
QY	481	caaggcagatgataaagccttataagacgctcgaaacccccatagaccttatccctacc	540
Db	1028	caaggcagatgataaagccttataagacgctcgaaacccccatagaccttatccctacc	1087
QY	541	tggctgattatgaacccctgatacgggtgtcaactttaaataagacactgaggtgtgtcctc	600
Db	1088	tggctgattatgaacccctgatacgggtgtcaactttaaataagacactgaggtgtgtcctc	1147
QY	601	agagagcacctggctggcctgaacatgcatcttcgcctccgattgatacccgctgttaa	660
Db	1148	agagagcacctggctggcctgaacatgcatcttcgcctccgattgatacccgctgttaa	1207
QY	661	agagacccctcccaacctatgcgtgaagtataggttctattgtcgtgccagagcacaga	720
Db	1208	agagacccctcccaacctatgcgtgaagtataggttctattgtcgtgccagagcacaga	1267
QY	721	gaagaaatactgtgggggtctcctgggaatccctctctctgaaggagatggagctgcgtca	780
Db	1268	gaagaaatactgtgggggtctcctgggaatccctctctctgaaggagatggagctgcgtca	1327
QY	781	aacgatgagacatggaatggccgcatctctccacagaaacgggtataaattcccttgc	840
Db	1328	aacgatgagacatggaatggccgcatctctccacagaaacgggtataaattcccttgc	1387
QY	841	aattccggcgccgggaagtataaataatgataaactatataaagataagatctgtcccca	900
Db	1388	aattccggcgccgggaagtataaataatgataaactatataaagataagatctgtcccca	1447
QY	901	tcaagactagattactaaagataaagtattcaacgaaagagaaacagaaataatctaa	960
Db	1448	tcaagactagattactaaagataaagtattcaacgaaagagaaacagaaataatctaa	1507
QY	961	aagtgatgaataatgctatagagctctggggaataagtttttaataatgacggggagcaagg	1020
Db	1508	aagtgatgaataatgctatagagctctggggaataagtttttaataataatgacggggagcaagg	1567
QY	1021	tccactttaaacatctgcgccttagagataagaaacgggagacaaacccctgttgcaatgga	1080
Db	1568	tccactttaaacatctgcgccttagagataagaaacgggagacaaacccctgttgcaatgga	1627
QY	1081	cccgataaagtaactgctcgtgaacaggggccccggcccttgagacacgcgcatataacttgcg	1140
Db	1628	cccgataaagtaactgctcgtgaacaggggccccggcccttgagacacgcgcatataacttgcg	1687
QY	1141	gtgcccgaataaactctgctgcgtgagcgtctgacataaaacagccgcctcagaaacagtaccact	1200
Db	1688	gtgcccgaataaactctgctgcgtgagcgtctgacataaaacagccgcctcagaaacagtaccact	1747
QY	1201	ggattgattcctacaaacacgccttagaagaatccccaagtggttccgtttaagaaacggagcag	1260
Db	1748	ggattgattcctacaaacacgccttagaagaatccccaagtggttccgtttaagaaacggagcag	1807
QY	1261	agactcttgctctcatccagggagcttcccaagccatcaaatccacacgacccctgatgac	1320
Db	1808	agactcttgctctcatccagggagcttcccaagccatcaaatccacacgacccctgatgac	1867
QY	1321	actctctctgttggtctgtcatcctcctaaggcctctttatitgagggagatgctaa	1380
Db	1868	actctctctgttggtctgtcatcctcctaaggcctctttatitgagggagatgctaa	1927
QY	1381	gaagaagaataatcaatgtgaccaaagagcatagaaatcaatgtacatgaggggtcccgaaat	1440
Db	1928	gaagaagaataatcaatgtgaccaaagagcatagaaatcaatgtacatgaggggtcccgaaat	1987

QY	1441	aaacttaacccctcaatgaagtttccgggagagggagcagcagtaagaaaatcccccattcc	1500
Db	1988	aagcttaacccctcaatgaagtttccgggagagggagacatgcaatgaagaaagctcccccattcc	2047
QY	1501	caaccacaaccttgcatactagtaactctgtgttttataatgagacagccctccagaataatcagatctta	1560
Db	2048	caaccacaaccttgcatactagtaactctgtgttttataatgagacagccctccagaataatcagatctta	2107
QY	1561	gtacacctgttatataaaggctgtgttgaggcaatgcatactctgtgttaacccccctgtttccaac	1620
Db	2108	gtacacctgttatataaaggctgtgttgaggcaatgcatactctgtgttaacccccctgtttccaac	2167
QY	1621	tcacgttcttaacccaatcccaaaagatctctgtgtcattgtgtcccaaatctgtcccccagctgtac	1680
Db	2168	tcacgttcttaacccaatcccaaaagatctctgtgtcattgtgtcccaaatctgtcccccagctgtac	2227
QY	1681	tacacatcctcgaggaagctgttcctttgataatgatactatcogytatataacccagaccaaaataga	1740
Db	2228	tacacatcctcgaggaagctgttcctttgataatgatactatcogytatataacccagaccaaaataga	2287
QY	1741	gaaccccgataccctcttaacccctagctgtataatgtctcgagtttgaggagccggtttggtgtatga	1800
Db	2288	gaaccccgataccctcttaacccctagctgtataatgtctcgagtttgaggagccggtttggtgtatga	2347
QY	1801	acaagaggaacagctgtccctctgatccaagagccacaagcacaactagaaagaagactctgtgtacata	1860
Db	2348	acaagaggaacagctgtccctctgatccaagagccacaagcacaactagaaagaagactctgtgtacata	2407
QY	1861	catgtgagccatgcagacaagaagaatctccgaagccttaaaagagctctgttagcaaccttaagaagag	1920
Db	2408	catgtgagccatgcagacaagaagaatctccgaagccttaaaagagctctgttagcaaccttaagaagag	2467
QY	1921	tcccctgaccttcttctgtctccgaagctgtgtcttcaacagaacccgagaggggttatgacctcttctt	1980
Db	2468	tcccctgaccttcttctgtctccgaagctgtgtgtcttcaacagaacccgagaggggttatgacctcttctt	2527
QY	1981	ctaagagaagaagctgtggtttatgtctcagaccttaaaagaagaatgttctgtcttaatgtatgatac	2040
Db	2528	ctaagagaagaagctgtggtttatgtctcagaccttaaaagaagaatgttctgtcttaatgtatgatac	2587
QY	2041	tcaagagacacatcaagaagactctcaatgcagaacagacttagaaaaaagttagagagagcgctgcgaag	2100
Db	2588	tcaagagacacatcaagaagactctcaatgcagaacagacttagaaaaaagttagagagagcgctgcgaag	2647
QY	2101	gaaagagaagcgtgaacccaagggctgtgttctgaaggaatgtgttccaacagctctccctgtgatac	2160
Db	2648	gaaagagaagcgtgaacccaagggctgtgttctgaaggaatgtgttccaacagctctccctgtgatac	2707
QY	2161	accctgaccttctgtctcgtgaaggggccccctagatgcctccgctccctgttaacttaacagtgtgg	2220
Db	2708	accctgaccttctgtctcgtgaaggggccccctagatgcctccgctccctgttaacttaacagtgtgg	2767
QY	2221	cctgtcttaataatgaagcttctgtctcctgttctlaagaagacagatgtgatgtgaagccagaatc	2280
Db	2768	cctgtcttaataatgaagcttctgtctcctgttctgttagagaaacagatgtgatgtgaagccagaatc	2827
QY	2281	atgtgtactttaggcacaagatcaacaagggcctctctgagccaaggaagaactgtacctctagacct	2340
Db	2828	atgtgtactttaggcacaagatcaacaagggcctctctgagccaaggaagaagaactgtacctcttagacct	2887
QY	2341	tcccagatctcaaaatattagaacattatcaaaacagaagaagctgtgggaataggaaaagtatgaata	2400
Db	2888	tcccagatctcaaaatattagaacattatcaaaacagaagaagctgtgggaataggaaaagtatgaata	2947
QY	2401	tgtcaacctaaacctccacagaacccagagaagttaataaaagctctataatgtcccccggaatt	2460
Db	2948	tgtcaacctaaacctccacagaacccagagaagttaataaaagctctataatgtcccccggaatt	3007
QY	2461	acagaacacctgtcgtctgcagaaataatgtatgaagaagctcaacacttctctattgttccaagggc	2520
Db	3008	acagaacacctgtcgtctgcagaaataatgtatgaagaagctcaacacttctctattgttccaagggc	3067
QY	2521	ctgtatactctgtgactaagtaaatataacagggaataatgtatgtactaatcogtatactctgtgatt	2580

```

|||||
Db 3068 cgtctactcctgctcaagaaagataacagaaatgagtgactaactcgttactcgtgalt 3127
Qy 2581 cgttaaactgactgacacacataagaagaattattacattgagacccctgtagcta 2640
Db 3128 ctgttaaacacgactgacacacataagaagaattattacattgagacccctgtagcta 3187
Qy 2641 tctcaactgcaatctgcaactctgcccagagagccacagacagatgagacccctgtagcta 2700
Db 3188 tctcaactgcaatctgcaactctgcccagagagccacagacagatgagacccctgtagcta 3247
Qy 2701 tttaaaatgattgtgtccacagagcgcgctcgtgataattttaaattgattgtgtgacatg 2760
Db 3248 tttaaaatgattgtgtccacagagcgcgctcgtgataattttaaattgattgtgtgacatg 3307
Qy 2761 gtagcgcgagctc-tcgatattttaaattgattgtgtgacagcagacgcttgtgtgta 2819
Db 3308 gtagcgcgagctcgtcgtatattttaaattgattgtgtgacagcagacgcttgtgtgta 3367
Qy 2820 acccacaataaagctgtcccgactccgacactcgaggcgccgacgtccttaccctcgctgtgt 2879
Db 3368 acccacaataaagctgtcccgactccgacactcgaggcgccgacgtccttaccctcgctgtgt 3427
Qy 2880 gtagcactgtgtggcccgacgcgctgtggaataaaatcctctgtctgtgtgacatc 2934
Db 3428 gtagcactgtgtggcccgacgcgctgtggaataaaatcctctgtctgtgtgacatc 3482

```

RESULT 2

US-08-766-528-2
Sequence 2, Application US/08766528
Patent No. 6190861

GENERAL INFORMATION:

APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 60 State Street
CITY: Boston
STATE: Massachusetts

COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766, 528
FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/572, 645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers
REGISTRATION NUMBER: 35, 965

REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 7333 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: CDNA
US-08-766-528-2

Query Match

27.5%; Score 813; DB 4; Length 7333;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 863; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 2082 gtagaagagcgctcgaagaggaagagagcgctgaccagaggtgtgttaagaattgttcaa 2141
Db 6470 GTTAGAGAGGCGGTGAAAGGAAAGAGAGCGGACGAGGCGGTGTTGAAGATGTTTCAA 6529
Qy 2142 caagctctccttgatgagacacccctgcttctgctctgacgagggcccttagtctgct 2201
Db 6530 CAGGTCCTCTTGATGATGACACCCCTGCTTTCCTCTGACGAGGCGCCCTAGTAGCTGCT 6589
Qy 2202 cctgttacttaacgttggcgcttgcttaataagttgttgccttgtttaagaagc 2261
Db 6590 CCTGTACTTACAGTTGGGCGCTTGCTTAAATATAGTTGTGCTTTGTTAAGAACG 6649
Qy 2262 agtggatgcaatccagatcatgattgattaggaagacagacacagccttctgagccaag 2321
Db 6650 AGTGAATGCAATGCCAGATCATGATGATTAAGCAACAGTACCAAGGCTTGTGAGCAAG 6709
Qy 2322 aagaactgacactagccttcccaqtlctaaattagaactatlaacaagaagaagtg 2381
Db 6710 AGAAGTGAACCTCTAGCCCTTCCAGTTCTAAGATTAGAATTAACAGAACAGACAGTG 6769
Qy 2382 gggaaatgaaagatgaaatgcaactaaccctcccaagacccaggaagttaataaag 2441
Db 6770 GGGAAATGAAGGATGAAATGCAACCTTACCTCCAGAACCCAGAGATTATATAAAG 6829
Qy 2442 ctctaattccccgaattacagacccctgtgctgacgagtaaatagtagaaggtcaca 2501
Db 6830 CTCTAATGCCCCCGAATTCAGACCCCTGCTGCTGCCAGTAATAGTAGTAAGGTACACA 6889
Qy 2502 ctctcattgttccagggcctgctacccctgacgaagtagaataacaggaattgattga 2561
Db 6890 CTTCCTATTGTTCCAGGCGCTGCTATCCCTAGCTAAGTAAACAGAAATGAGTTGA 6949
Qy 2562 ctaactgctatctgattctgttaaacctgactgagacataagaagaattgattacacat 2621
Db 6950 CTAATGCTTATCTGATTTCTGTAAAGCTGACTGACACCATAGAGAAATTGATTACACT 7009
Qy 2622 tgacagccctagtgactctatctcaactgcaatctgcaactctgcccagagccacgacg 2681
Db 7010 TGACAGCCCTAGTGACTCTATCTCAACTGCAATCTGTCACTCTGCCAGAGCCACGACG 7069
Qy 2682 atgcggaactcggagctattttaaattgattgtgtccagggcgcgctcgtgattt 2741
Db 7070 ATGCGGACCTCCGAGCTATTATTAATGATTTGTCACGAGCGGCGCTCTCGATATT 7129
Qy 2742 ttaaaatgattgttccatgagagcgcgctcgtgataattttaaattgattgtgtgac 2801
Db 7130 TTAATAATGATTGTTCATGAGAGCGCGGCTCTGATATTTTAATAATGATTGCTTGTGAC 7189
Qy 2802 gcaacagccttgtgtgaacccataaaagctgtcccgattccgcaactcgaggcgagtg 2861
Db 7190 GCACAGGCTTGTGTTGTAACCCCATATAAAGCTGTCCGATCCGCACTCGGCGCCAGT 7249
Qy 2862 cctctacccctggtgtgtgaagactgtgtggccacagcgactgttgaaataaactcctc 2921
Db 7250 CCTCTACCCCTGCGCTGTGTGACACTGTGTGGCCCAAGCGCTTGGATTAATAATCTCT 7309
Qy 2922 tgcgtgttgatcaaaaaaataa 2945
Db 7310 TGCTGTTTGCATCAAAAAAATAA 7333

```

RESULT 3

US-09-111-085-1
Sequence 1, Application US/09111085
Patent No. 6100034

GENERAL INFORMATION:

APPLICANT: Stoye, Jonathan P
APPLICANT: Weiss, Robin A
TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
TITLE OF INVENTION: specific sequences

FILE REFERENCE: 4238/75168
 : CURRENT APPLICATION NUMBER: US/09/111,085
 : CURRENT FILING DATE: 1998-07-07
 : EARLIER APPLICATION NUMBER: GB 9710154.7
 : EARLIER FILING DATE: 1997-05-16
 : NUMBER OF SEQ ID NOS: 16
 : SOFTWARE: Patentln Ver. 2.0
 : SEQ ID NO 1
 : LENGTH: 2462
 : TYPE: DNA
 : ORGANISM: Porcine retrovirus
 : US-09-111-085-1

Query Match 15.5%; Score 457; DB 3; Length 2462;
 Best Local Similarity 99.8%; Pred. No. 4,6e-195;
 Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2082 gttagaagagcgtcgaaggaagagagctgaccaggggtgtgttgaagatgttcaa 2141
 Db 1938 gttagaagagcgtcgaaggaagagagctgaccaggggtgtgttgaagatgttcaa 1997
 QY 2142 caaggtcctctggtatgaccacccgtgtctgtctgacgggggccccctagttctctgt 2201
 Db 1998 caaggtcctctggtatgaccacccgtgtctgtctgacgggggccccctagttctctgt 2057
 QY 2202 cctgttacttaccagttgggcttgccttaataatagttgttgccttctgttagaagacg 2261
 Db 2058 cctgttacttaccagttgggcttgccttgaataatagttgttgccttctgttagaagacg 2117
 QY 2262 agtgaatgacgtcagacacatgcttacttaggaagacagtaaccagccctcgtgaccaaag 2321
 Db 2118 agtgaatgacgtcagacacatgcttacttaggaagacagtaaccagccctcgtgaccaaag 2177
 QY 2322 agaactgacctgacctcctccagttccttaagattagactataacaaagaaagtg 2381
 Db 2178 agaactgacctgacctcctccagttccttaagattagactataacaaagaaagtg 2237
 QY 2382 ggggaatgaaagatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 2441
 Db 2238 ggggaatgaaagatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 2297
 QY 2442 cttctaaatgccccgaattacagacccgtgctgctgacgaataatgtagaagttaca 2501
 Db 2298 cttctaaatgccccgaattacagacccgtgctgctgacgaataatgtagaagttaca 2357
 QY 2502 cttctaatgttccagggcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2561
 Db 2358 cttctaatgttccagggcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2417
 QY 2562 ctaacgtctatctgattctgttaaac 2589
 Db 2418 ctaacgtctatctgattctgttaaac 2445

RESULT 4
 : US-09-376-781-8
 : Sequence 8, Application US/09376781
 : Patent No. 6261806
 : GENERAL INFORMATION:
 : APPLICANT: Banerjee, Papi T.
 : APPLICANT: Patience, Clive
 : APPLICANT: Andersson, Goran K.
 : TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
 : Patent No. 6261806
 : FILE REFERENCE: 61750-267
 : CURRENT APPLICATION NUMBER: US/09/376,781
 : CURRENT FILING DATE: 1999-08-18
 : EARLIER APPLICATION NUMBER: 60/097,015
 : EARLIER FILING DATE: 1998-08-18
 : NUMBER OF SEQ ID NOS: 33
 : SOFTWARE: Patentln Ver. 2.0

: SEQ ID NO 8
 : LENGTH: 276
 : TYPE: DNA
 : ORGANISM: Artificial Sequence
 : FEATURE:
 : OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
 : OTHER INFORMATION: sequence of part of PERV-B env gene corresponding
 : OTHER INFORMATION: to residues 1350-1623 of Genbank Accession No. 6261806
 : OTHER INFORMATION: Y12239.
 : US-09-376-781-8

Query Match 5.0%; Score 147; DB 4; Length 276;
 Best Local Similarity 99.5%; Pred. No. 1.7e-56;
 Matches 197; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 801 gccgattctctccagagccgggttaaatctcccttctgcaattccggccgggcaagta 860
 Db 1 gccgattctctccagagccgggttaaatctcccttctgcaattccggccgggcaagta 60
 QY 861 caaatgataaactataataaagataagagctgctcccatcagactatatactaaa 920
 Db 61 caaatgataaactataataaagataagagctgctcccatcagactatatactaaa 120
 QY 921 gataagttcactgaaagaaagaaacaggaataatttaaaagtggaataatgtatgag 980
 Db 121 gataagttcactgaaagaaagaaacaggaataatttaaaagtggaataatgtatgag 180
 QY 981 ctgggaatagttttta 998
 Db 181 ctgggaatagttttta 198

RESULT 5
 : US-09-376-781-10
 : Sequence 10, Application US/09376781
 : Patent No. 6261806
 : GENERAL INFORMATION:
 : APPLICANT: Banerjee, Papi T.
 : APPLICANT: Patience, Clive
 : APPLICANT: Andersson, Goran K.
 : TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
 : Patent No. 6261806
 : FILE REFERENCE: 61750-267
 : CURRENT APPLICATION NUMBER: US/09/376,781
 : CURRENT FILING DATE: 1999-08-18
 : EARLIER APPLICATION NUMBER: 60/097,015
 : EARLIER FILING DATE: 1998-08-18
 : NUMBER OF SEQ ID NOS: 33
 : SOFTWARE: Patentln Ver. 2.0
 : SEQ ID NO 10
 : LENGTH: 278
 : TYPE: DNA
 : ORGANISM: Artificial Sequence
 : FEATURE:
 : OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
 : OTHER INFORMATION: sequence of part of PERV-B env gene corresponding to
 : OTHER INFORMATION: residues 6057-6333 of WO 97/40167 for comparison.
 : US-09-376-781-10

Query Match 4.6%; Score 137; DB 4; Length 278;
 Best Local Similarity 100.0%; Pred. No. 5.1e-52;
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 801 gccgattctctccagagccgggttaaatctcccttctgcaattccggccgggcaagta 860
 Db 1 gccgattctctccagagccgggttaaatctcccttctgcaattccggccgggcaagta 60
 QY 861 caaatgataaactataataaagataagagctgctcccatcagactatatactaaa 920
 Db 61 caaatgataaactataataaagataagagctgctcccatcagactatatactaaa 120

OY 921 gataagttcactgaa 937
|||||
Db 121 gataagttcactgaa 137

RESULT 6

US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-3

Query Match 3.3%; Score 98; DB 4; Length 8132;
Best Local Similarity 100.0%; Pred. No. 1.1e-34;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2846 caccggggccgagctcttaccctcctggtgtgaagctgtggcccgagcgctt 2905
|||||
Db 8035 cactcggggccgagctcttaccctcctggtgtgaagctgtggcccgagcgctt 8094
OY 2906 ggaataaaatcctctgtctgtgtgcatcaaaaaa 2943
|||||
Db 8095 ggaataaaatcctctgtctgtgtgcatcaaaaaa 8132

RESULT 7

US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match 3.1%; Score 91; DB 4; Length 8060;
Best Local Similarity 100.0%; Pred. No. 1.5e-31;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2846 cactcggggccgagctcttaccctcctggtgtgtgaagctgtggcccgagcgctt 2905
|||||
Db 2501 cactcggggccgagctcttaccctcctggtgtgtgaagctgtggcccgagcgctt 2560
OY 2906 ggaataaaatcctctgtctgtgtgcatcaaa 2936
|||||
Db 2561 ggaataaaatcctctgtctgtgtgcatcaaa 2591

RESULT 8

US-09-376-781-24
; Sequence 24, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; EARLIER FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
; OTHER INFORMATION: Sequence of the 3' end of the PRV-D env region.
US-09-376-781-24

Query Match 0.9%; Score 28; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2047 gccatcagagactccatgaacaagctta 2074
|||||
Db 1135 gccatcagagactccatgaacaagctta 1162

RESULT 9

US-09-360-197-13
; Sequence 13, Application US/09360197
; Patent No. 6287859
; GENERAL INFORMATION:
; APPLICANT: Bassilana, Frederic
; APPLICANT: Lazdunski, Michel
; APPLICANT: Waldmann, Rainer
; APPLICANT: Deweille, Jan R.
; TITLE OF INVENTION: Human and Rat Families of Neuronal Acid-Sensitive
; FILE REFERENCE: 989.6706P
; CURRENT APPLICATION NUMBER: US/09/360,197
; CURRENT FILING DATE: 1997-07-23
; PRIOR APPLICATION NUMBER: 09/129,758
; PRIOR FILING DATE: 1998-08-05
; PRIOR APPLICATION NUMBER: 60/095,408
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-360-197-13

Query Match 0.9%; Score 28; DB 4; Length 1736;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2929 tgcataaaaaaaaaaaaaaaaaa 2956
|||||
Db 1709 tgcataaaaaaaaaaaaaaaaaa 1736

RESULT 10

US-09-376-781-25
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES: Description of Artificial Sequence:Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOS: 18, 22, and 24.
US-09-376-781-25

Query Match 0.9%; Score 28; DB 4; Length 2000;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2047 gccatcagagactccatgaacaagctta 2074
|||||
Db 1654 gccatcagagactccatgaacaagctta 1681

RESULT 11

US-09-376-781-30
; Sequence 30, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 30
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: consensus sequence of PERV-D.
US-09-376-781-30

Query Match 0.9%; Score 28; DB 4; Length 2000;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2047 gccatcagagactccatgaacaagctta 2074
|||||
Db 1654 gccatcagagactccatgaacaagctta 1681

RESULT 12

US-09-108-020-11
; Sequence 11, Application US/09108020A
; Patent No. 6143561
; GENERAL INFORMATION:
; APPLICANT: Randall, Douglas D.
; APPLICANT: Johnston, Mark L.
; APPLICANT: Miernyk, Jan A.
; APPLICANT: Luethy, Michael H.
; TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
; TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
; TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANATE BIOSYNTHESIS IN PLANTS
; FILE REFERENCE: UMO 1482
; CURRENT APPLICATION NUMBER: US/09/108,020A
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 60/051,291
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 60/055,255
; EARLIER FILING DATE: 1997-08-01
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/076,544
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 11

LENGTH: 1587
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-108-020-11

Query Match
Best Local Similarity 100.0%; Score 27; DB 3; Length 1587;
Pred. No. 0.0065;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2930 gcatcaaaaaaaaaaaaaaaaaa 2956
|||||
DB 1553 gcatcaaaaaaaaaaaaaaaaaa 1579

RESULT 13
US-08-831-570-2/c
Sequence 2, Application US/08831570
Patent No. 5959175

GENERAL INFORMATION:
APPLICANT: Thomas, Terry L.
APPLICANT: Nuber, Andrew N.
APPLICANT: Beremand, Phillip D.
TITLE OF INVENTION: A SUNFLOWER ALBUMIN 5' REGULATORY REGION
TITLE OF INVENTION: FOR THE MODIFICATION OF PLANT SEED LIPID
NUMBER OF INVENTIONS: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,570
FILING DATE: 09-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DiGioglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10545
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 743-4366
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 843 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-831-570-2

Query Match
Best Local Similarity 100.0%; Score 26; DB 2; Length 843;
Pred. No. 0.019;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2931 catcaaaaaaaaaaaaaaaaaa 2956
|||||
DB 305 CATCAAAAAAAAAAAAAAAAAA 280

RESULT 14
US-09-248-335-35
Sequence 35, Application US/09248335
Patent No. 6096504

GENERAL INFORMATION:
APPLICANT: MCSONIGLE, BRIAN
APPLICANT: O'KEEF, DANIEL
TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1128-A
CURRENT APPLICATION NUMBER: US/09/248,335
CURRENT FILING DATE: 1999-02-10
EARLIER APPLICATION NUMBER: 08/924,759
EARLIER FILING DATE: 1997-September-05
NUMBER OF SEQ ID NOS: 74
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 35
LENGTH: 1098
TYPE: DNA
ORGANISM: maize
US-09-248-335-35

Query Match
Best Local Similarity 100.0%; Score 26; DB 3; Length 1098;
Pred. No. 0.019;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2931 catcaaaaaaaaaaaaaaaaaa 2956
|||||
DB 1006 catcaaaaaaaaaaaaaaaaaa 1031

RESULT 15
US-08-725-532A-2/c
Sequence 2, Application US/08725532A
Patent No. 6020179
GENERAL INFORMATION:
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,532A
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0130 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-725-532A-2

Query Match 0.9%; Score 26; DB 3; Length 1275;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2931 catcaaaaaaaaaaaaaaaaaa 2956
|||||
Db 40 CATCAAAAAAAAAAAAAAAAAA 15

Search completed: February 24, 2002, 07:31:17
Job time: 16486 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 21:04:09 ; Search time 212.72 seconds
(without alignments)
3534.726 Million cell updates/sec

Title: US-09-171-553B-1
Perfect score: 3320
Sequence: 1 gaattcgccgcgcgcgtgcac.....aaaaaaaaaaaaaaaaaaaaa 3320

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTC05.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2915.2	87.8	8132	4	US-08-766-528-3
2	2420.6	72.9	8060	4	US-08-766-528-1
3	1771.8	53.4	7333	4	US-08-766-528-2
4	1175.8	35.4	8655	3	US-09-075-272-1
5	1131	34.1	3482	3	US-09-111-065-3
6	1088	32.8	8535	3	US-08-716-351A-1
7	933.6	28.1	8323	1	US-08-110-300A-8
8	933.6	28.1	8323	2	US-08-886-642-8
9	933.6	28.1	8323	5	PCR-US93-08041-8
10	932	28.1	10367	1	US-08-110-300A-9
11	932	28.1	10367	2	US-08-886-642-9
12	932	28.1	10367	3	PCR-US93-08041-9
13	916.2	27.6	7308	4	US-09-011-745-3
14	916.2	27.6	7308	4	US-09-011-745-4
15	916.2	27.6	7308	4	US-09-011-745-2
16	916.2	27.6	7308	4	US-09-011-745-1
17	910.8	27.4	8332	3	US-08-850-961-1
18	847.8	25.5	2462	1	US-08-258-420-13
19	624.6	18.8	6363	2	US-09-111-085-1
20	423.8	12.8	445	4	US-09-376-781-7
21	372.4	11.2	698	4	US-09-376-781-18
22	371.8	11.2	2000	4	US-09-376-781-25
23	371.8	11.2	2000	4	US-09-376-781-30
24	307.2	9.3	1140	4	US-09-603-185-5
25	305.2	9.0	6312	4	US-09-011-745-7
26	300.4	9.0	949	5	PCR-US91-08254-1
27	300.4	9.0	949	5	PCR-US91-08254-2

28	233	7.0	735	1	US-08-798-000-3	Sequence 3, Appl1
29	226.6	6.8	1234	1	US-08-798-000-1	Sequence 1, Appl1
30	221	6.7	1237	1	US-08-798-000-2	Sequence 2, Appl1
31	218	6.6	3674	1	US-08-105-483-324	Sequence 324, App
32	218	6.6	3674	1	US-08-709-209-324	Sequence 324, App
33	218	6.6	3674	1	US-08-458-101-324	Sequence 324, App
34	216	6.5	9388	4	US-08-991-789A-141	Sequence 141, App
35	204	6.1	3646	4	US-08-991-789A-227	Sequence 227, App
36	183.4	5.5	2518	4	US-09-011-745-1	Sequence 1, Appl1
37	165.6	5.0	2448	3	US-08-691-563C-53	Sequence 53, Appl1
38	149.8	4.5	1167	3	US-08-691-563C-61	Sequence 61, Appl1
39	141.2	4.3	2389	3	US-08-691-563C-52	Sequence 52, Appl1
40	134.4	4.0	1101	4	US-08-991-789A-3	Sequence 3, Appl1
41	119.2	3.6	1158	1	US-08-471-724-1	Sequence 1, Appl1
42	119.2	3.6	1158	2	US-08-471-969-1	Sequence 1, Appl1
43	119.2	3.6	1158	2	US-08-384-137-1	Sequence 1, Appl1
44	119.2	3.6	1158	2	US-08-470-006A-1	Sequence 1, Appl1
45	119.2	3.6	1158	3	US-08-691-563C-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-3
Query Match 87.8%; Score 2915.2; DB 4; Length 8132;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 3051; Conservative 0; Mismatches 173; Indels 12; Gaps 2;
22 gatctctcttcgtgctgagattacacccacatagccaacacttttgcctcgatgg 81
|||||

Dh 2970 GATGCTTCTTTCGCTGAGATTACACCCACTAGCCAAACACTTTTTCGCTTCGANTGG 3029
Qy 82 agagatccaggtacggaagaaacgggcaagctcaactgagaccgagctgcccagaaggttc 141
Dh 3030 AGAGATCCAGGTACGGGGAAGAACCGGGAGCTCAGCTGAGCCGAGCTGCCCCAAGGGTTC 3089
Qy 142 aagaatcccccagccatcttgaagagccctacacagagaccctggccaactcagatc 201
Dh 3090 AAGAATCCCCGACCACTTTTGACGAAAGCCCTACACAGAGACCTGGCCCAACTTCAGGATC 3149
Qy 202 caaacaccccaagtgagccctccctccagtgagtgagagaccctctcttgccggaagccac 261
Dh 3150 CAACACCCCTCAGTGAGCTCCTCCTCCAGTACGTGATGACCTCTCTTGCGGGAGCCACC 3209
Qy 262 aaacagagactgcttagaaggtacgaagagcactactgctgaaattgctgaacctagctac 321
Dh 3210 AAACAGAGACTGCTTAGAAGGACGACGAGGACTACTGCTGGAGTTGCTGACCTTAGGCTAC 3269
Qy 322 aagaacctgcttagaagaaagccagatttgcaagagagaggtacaacttggggtacagt 381
Dh 3270 AAGACCTCTGCTTAGAAGAGCCGACGATTGTCAGAGAGAGGTAACTACTTGGCGTACGT 3329
Qy 382 ttgcgagggcgagcagatgctgagcagagcaggaagaaactgtagtccagatcagc 441
Dh 3330 TTGCGGAGCGGGCAGCGAGTGGCTGACGAGGACGGAAGAAACTGTAGTCCAGATACCG 3389
Qy 442 gccccaacacagacgaacaaagtgagagagttttggggaagcgtgtgaatttgcagactg 501
Dh 3390 GCCCAACACACGACCAAACTAGAGAGTTTGGGGAGACGCTGAGTTTGGAGACTG 3449
Qy 502 tggatcccgagggtttgagaccttagcagccacactctcccgcttaccacaaagaaaggg 561
Dh 3450 TGGATCCCGGGGTTTGGCACTTACAGACCCCACTACCTGCGCTAACCAAGAAAAAGGG 3509
Qy 562 ggaattccctgggctcctgtagacacacgaagagcatttgaatgctatcaaaagggccctgtg 621
Dh 3510 GAATTCCTCTGGGCTCTGAGACACGAGAGGATTTGATGCTATCAAAAAGGCGCTGCTG 3569
Qy 622 agcgagactgctctgagccctccctgagcgttaactaaaccttacccttattgtgagtacg 681
Dh 3570 AGCGACCTGCTGCGCCCTCCCTGACACTAAACCCCTTACCTTTATGTGGATGAG 3629
Qy 682 cgtagaggagtagcccgagaggttttaacccaacccctagacatagagagagccgtt 741
Dh 3630 CGTAGAGGAGTAGCCCGGGGAGTTTAAACCAAAACCTAGGACCATGAGAGAACCTGTC 3689
Qy 742 gctcaactgtcaaaagagcttgatccctgtagccagtggttgagccgtatgctaaagct 801
Dh 3690 GCTCACTGTCAAAAGAACTCGATCCTGTAGCCAGTGGTGGCCCATATGCTGAAAGCT 3749
Qy 802 atcgagctgtgagcactactgtgtaagagagcgtcaaaattgactttgggaacagatatata 861
Dh 3750 ATCGAGCTGTGGCCATCTAGTGTCAAGGAGCCTGACAAATTTGAGCTTTGGGACAAATATA 3809
Qy 862 actgtaataagcccccaatgcatgtgagaacatcgctgcgacgcccccaagacccaatgagt 921
Dh 3810 ACTGTAATAGCCCCCATGATGTGAGAGAACTGTTGGCGACGCCCCAGACGCGATGAGT 3869
Qy 922 accaagcccgcatgacccactatcaaaagcgtcttctcaagaagaggtcaacttgcgt 981
Dh 3870 ACCAAGCCCGCATGACCCACTATCAAAAGCTGCTTCTCAAGAGAGAGGTACAGTTGCT 3929
Qy 982 ccaccagcgagctcaaacctcgacacactctctgctgtaagagagactgtaacaaagtact 1041
Dh 3930 CCACACAGCGGCTTCAACCCGCGACCTTCTGCTGAAAGACTGTATGAAACACAGTACT 3989
Qy 1042 catgattgccatcaactattgattgagagagagctgggttcgcaagagaccttaagacata 1101
Dh 3990 CATGATTGCCATCACTATTGATTGAGAGAGCTGGGTCCGCAAGGACCTTAAAGACATA 4049
Qy 1102 ccgctgactggaagagtgtaacctggttcaactgagcgaagcagctatggtgtggaaggt 1161
Dh 4050 CCGCTGACTGGAGAAAGTGTACCTGTTCACTAGCAGGAGCACTATGTGTGGTGAAGGT 4109

Qy 1162 aagaagatgctggggcgagtggtgagcaggaaccccgacgacttgggcccagagcctg 1221
Dh 4110 AAGAGAGATGCTGGGGCGCGCGGTGTGACGGGACCCGACGATCTGGCGACAGCTCT 4169
Qy 1222 ccggaaggaactcagcgcaaaagcctgagctcaatggccctcaacgcaagcttggcgctg 1281
Dh 4170 CCGGAGGAACTTACGACCAAAAGGCTGAGTCTATGGCCCTCAACGCAAGCTTGGCGCTG 4229
Qy 1282 gccgaagggaaatccataaacttatacggagacagcagatagccttgcgactgcaac 1341
Dh 4230 GCCGAGGGAATCCATTAACATTTATACGACGACGAGTATGCTTGGAGCTGCACAC 4289
Qy 1342 gtacacgggccatctataacaaaggggttgcctacccccaagggaggaagaaataag 1401
Dh 4290 GTACATGGGGCATCTTAAACAAAGGGGTTGCTTACCTCAGCAGGAGGAAATTAAG 4349
Qy 1402 aacaaagaggaatcttaagcctttagaagccttaccattgccaataaagcctagctat 1461
Dh 4350 AACAAAGAGAAATTTTAAGCCTATTAGAACCGGTACATTACCAAAAAGCGTACGCTAT 4409
Qy 1462 atacactgtcctgagactcaagaagcccaagatcctatctagagggaacagatgct 1521
Dh 4410 ATACACTGCTCTGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATGCT 4469
Qy 1522 gaccgggttgccaagcagagcccgagcgtcttacccttgcctataatagaaagccc 1581
Dh 4470 GACCGGGTTGCCAAGCAGCGAGCCAGCGTGTAACTTGTGCTTATATAGAAATGCC 4529
Qy 1582 aaagccccaagaaacccaagcagatcacaccttagaagacttgcgaagagataaaaaagata 1641
Dh 4530 AAAGCCCCAGAACCCAGACAGCAGTACACCTTGAAGAGCTGCAAGAGATTAATAAAGATA 4589
Qy 1642 gaccagttctctgagctccggaggagacctgttatactcataatgggaagaaatccgt 1701
Dh 4590 GACCAAGTCTCTGAGACTCCGGAAGGACTGCTATACCTCAATGGAAGAAATCCGT 4649
Qy 1702 ccccaacaaagaaaggttagaatatgtccaacagatcatcgttcaacccaactagaagact 1761
Dh 4650 CCCCAACAAAGAGGTTAGAAATGTGTCCAAACATATCATGCTTAACCCACCTAGAACT 4709
Qy 1762 aaacactgacagagtggttgcagaaatcccttcatcattgttgcagctaccagagtg 1821
Dh 4710 AAACACCTTGACGAGATGTGTCAGAACATCCCTTATCATGTTCTGAGCTACAGAGAGT 4769
Qy 1822 gctgactcgtgtgtcaaacatgtgtgccttgcacagctggttaagtctatccctcaga 1881
Dh 4770 GCTGACTCGGTGTCAAACTTTGTGTGCTTGCACAGCTGTTAATGCTAATCCTTCCAA 4829
Qy 1882 atacctccagaaagagactaaaggggaagccacccagcgctcacttgggaagtggacttc 1941
Dh 4830 ATGCTTCCAGGAAAGAGACTAAAGGGGAGACCAACCCAGGCGTCACTGGGAATGAGCTTC 4889
Qy 1942 actgagtagtaagccgcttaatacggaaacaaatatctatgtgtttgttagaacctt 2001
Dh 4890 ACTGAGGTAAAGCGGCTTAATAACGGAACCAATACCTATTGTTGTTGTAGACACTTT 4949
Qy 2002 tcaagga tgggtagaagctatccctactaaagaaagagacttcaacggttggcttagaa 2061
Dh 4950 TCAAGGATGGGTAGAGGCTTATCTTACTAAGAAAGAGACTTCAACCGTGTGCTAAAAA 5009
Qy 2062 atactgaggaatatttccaaagatttgaatacctaaaggtlaataggttcaagacaatggt 2121
Dh 5010 ATACTGGAAGAAATTTTCCAAAGATTGGAATACCTAAGGTAAATAGGTGACAGACATGCT 5069
Qy 2122 cgaagcttgcgttgcgaaggttaagtcagagagctgccaagatatttgggagattgtgaaa 2181
Dh 5070 CGAGCTTTGTTGGCCGAGTAAAGTCAAGGACGTGCGCAAGATATTGGGATGATTGAAA 5129
Qy 2182 ctgcatgtgagctacagaccccaaggtccagagacaggttagagagagataagaaacact 2241
Dh 5130 CTGCAATTGTGATACGAGACCCCAAGACTCAGAGACAGGTAGAGAGATGAATGAACCAT 5189

QY	2242	aagaagacccttaaccaaatgtgacacagagactgcaatlaaagatlgagtgctctccg	2301
Db	5190	AAAGAGACCCCTTAACATAATGACCCCGAGACTGGCTTAATGATGATGATACCTCTCCG	5249
QY	2302	cccttctgctttttagaggtgaggaacaccccttgacagcttggcttgacccctatgaa	2361
Db	5250	CCCTTTTGCTTTTTAGGGTTAGGAACACCCCTTGACAGTTTGGCTGACCCCTATGAA	5309
QY	2362	tgtctcaagggagacccccccgcttggcagaatgtcctttgcaatgctgctatg	2421
Db	5310	TTACTTACAGGGGAGACCCCCCATTTGATGAATAATTCCTTGTACATATGCTGACGTG	5369
QY	2422	ctgcttccccagccttgtctctcagctcagctcaagcgctcgagtggggtgagcgagcg	2481
Db	5370	CTGCTTCCAGCCCTTGTGTTCTTAGCTCAAGGAGACTTGAGTGGGGTAGACACAGAGCG	5429
QY	2482	tggaagacagctccggagagagcctactc--agagagagactgtcaagttcacatgccttc	2538
Db	5430	TGGAGGCAACTCCGGGAGGCTACTCAGAGGAGGAGAGATCTGCAGATCCACATTCGTTTC	5489
QY	2539	caagttgagatctcagctctatgtgtgagacgcaacggtgagaanaacctcagactcgg	2598
Db	5490	CAAGTGGAGATTCAGTTCAGTTCAGTTCAGACCCACCGCTGCAGGAACCTCGAGACTCGGTG	5549
QY	2599	aaggaaccttctcgcactcttgacacacacacaagcgctgtgaaagtcgagaagatcccc	2658
Db	5550	AAGGGCCCTTATCTCGTATCTTTTGACACACACCAAGCGCTGGAAGAATCTCC	5609
QY	2659	acctgcatcatgcatcccaacgcttaagcgcgcgcaactcccgcttgggggttggaaagcc	2718
Db	5610	ACCTGGATTCATGATGCCACGTTAAACGGGGCCACCGCTCCGATTCGGGGTGGAAAGCC	5669
QY	2719	gaaagactggaataccoccttaagctcgcctccatcgcgctggttcccttactctgtcaat	2778
Db	5670	GAAAGACTGAAATCCOCTTAAGCTTCGGCTTCATCGCGTGTCTCTTACTCTGTCAAT	5729
QY	2779	aactcctcaagttaatgtlaaagccttgtggaacgccccgaactccctcaaaccttatc	2838
Db	5730	AACCTCTCAGATAATGTTAGTACGCCATGAGAGACACCGTGAACCTCCATTAACCTTATC	5789
QY	2839	tctcaccttgcttacttacttgactcgcggtacagatataatlaatacagcactcaagggga	2898
Db	5790	TCTCAGCTTGTTAATTACTGACTCGGGCACAGGTATTATATCAACAACACTCAAGGGGA	5849
QY	2899	ggtctcccttgggagaccggtggtgcctgaattatatatgctgctcttgactgaagtaaccgtg	2958
Db	5850	GGCTCCTTAGAGAACCTGCTGCTGATATACGTTTACGTTTGCTCGAATCAAGTTATCTCAG	5909
QY	2959	tctcaatgacacgagcacaaccccccgatgtaactcgtgcgtcagcaggtttacgtttgcc	3018
Db	5910	TCT-----GACCTCACCCCCAGATATCTTCATAGCTCAAGGATTTATGTTTGCC	5960
QY	3019	aggaccccccaataatgaaagatattgtgaaacatccatcgaagattcccttggaaagcaatg	3078
Db	5961	AGGACCAACCAATAATGAGAAACATTTGGGGAATCCCAAGATTTTGTGTAACAATG	6020
QY	3079	gagctgcataactctctaagatagggaattgnaatgtgcacgctctcacaagaagaagat	3138
Db	6021	GAACCTGTATACCTCTTAATGATGATGATTTGGAATTTGGCAACCTCTCACAGAGATAGGT	6080
QY	3139	aagttactcttctgttaaacatccatccacagttacatatcaatlaataatgagccatgggag	3198
Db	6081	AAGTTTCTTATGATCAACACTATACAGCTCTGCAACATTTAATTTACCTGACCTGGAT	6140
QY	3199	atggaagaattgtgcaacacggtgacaaaagatgtaacgaataatgaagataagct	3254
Db	6141	TAGAACTGGAAGCCCCAAGTGTCTCTTCAGACTAGATTACTTAATAATATAGTT	6196

GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MCP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match	72.9%	Score 2420.6	DB 4	Length 8060
Best Local Similarity	97.5%	Pred. No. 0		
Matches 2502	Conservative	Mismatches	59	Indels 6
		Gaps		
QY 22	gattccctcttcgagctgagatatacaccacatagcaacacatttctgcttcgaatg	81		
Db 5495	GATGCTTTCTTTCGCTGAGATATACCCCTATACCAACACTTTTGCTTCGATGG	5554		
QY 82	agagatccagtlacgggaagaacccgggcgcctcacttgaccgcagatgcccagaagtgc	141		
Db 5555	AGAGTCTCAGGTACGGGAAGAACCCGGCAGCTTACTTGACCCGACTGCCAAGGTTG	5614		
QY 142	agaactccccgacacatctttagcagaagccctacacaagagacttggccaattgaatc	201		
Db 5615	AAGACTCCCGGACCATCTTTGACGAAGCCCTTACAGAGACTCTGGCCACTTGAGATC	5674		
QY 202	caaacacctcaagtgtaccctccctccgtagtcgtgagatgcctctcttgcgggagacc	261		
Db 5675	CACACCTCTCAGGAGACCTCTCTCCAGTACGTGATGACTCTTCTGGCGGAGCCAC	5734		
QY 262	aaacagagacttcttagaagtlacgaagcaactactctggaattgtctgcacttggtctac	321		
Db 5735	AAMCAGGACTGTTTAGAAGGCACGAAGGCACTACTCTGGAATTTCTGTGACCTTAGGCTAC	5794		
QY 322	agagaccttgcttaagaagagcccaatttgaagaagaagatataactttgggtacagt	381		
Db 5795	AGAGCCTCTGCTAAAGAGGCCCAATTTGACAGAGAGAGTAACTACTTGGGGTACAGT	5854		
QY 382	tttcggcgcgagcgacgatlgtctlgacggagcaacggaagaacgttagtcagaatacc	441		
Db 5855	TTTACGGGACGGCACGTAGCTGACGAGAGCAACGGAAGAAACTTATCTCAGATACCG	5914		
QY 442	gccccaaccacacgaacaaacagtgagagagtttttgggacagcttgattttgcagactg	501		

Db 5915 GCGCCCAACCAAGCCAAACAAATGAGAGAGTTTGGGGAAGCTGATTTTGGACAGCTG 5974
OY 502 Tgataccggaggttgagacaccttagcagcccccactctaccccgctaaccaaaagaaaggg 561
Db 5975 TGGATCCCGGGGTTTGCACTTACACACCCACCTACCTACCTACCTACCTACCTACCTACCT 6034
OY 562 ggaattccctggcctcctgagacacagagacatttgatgatacaaaaagccctgctg 621
Db 6035 GAATTCCTCTGGGCTCTGAGACACCAAGAGCATTTGATGCTATCAAAAAAGGCCCTGCTG 6094
OY 622 agcgaactgctctgagccctccctgagcgttaacttaaaccttaaccccttaagtgtgagtg 681
Db 6095 AGCGACCTGCTCTGGGCTCTGAGACACCAAGAGCATTTGATGCTATCAAAAAAGGCCCTGCTG 6154
OY 682 cgttaaggagtagagcccgagaggttttaacccaaccccttagaccctgagagagagacctgt 741
Db 6155 CGTAAGGAGTAGAGCCCGGGAGGTTTAAACCAACCCCTAGACCAAGAGAGAGAGAGAGAG 6214
OY 742 gctcaactgtcaaaagagcttgatcctgtagccagtggttgagccgtatgcttgagagct 801
Db 6215 GCGTACCTGTCAAAGAGCTCGATCCGTGAGCCAGTGCTGAGCCATATGCTGAGAGCT 6274
OY 802 atcgagcgttgagccactgagcagagagcgtgacaaatgactttgagac-agaatat 860
Db 6275 ATGCGACCTTGCGCATACTGATGATAGAGAGCGCTGACAAATGACTTTGGACAAAGATAT 6334
OY 861 aactgtataagcccccatgacttgtagagacatcgttgcgagagccccagagagacctgat 920
Db 6335 AACTGTATATGCCCCCATGACTTGAGAGAAATCGCTTCGCGAGCCCAAGACCCGATGAT 6394
OY 921 gaccaagcccgagtagaccactatacaaaagcctgcttctcaagagagagcttgcac 980
Db 6395 GACCAAGCCCGCATGACCCCATATCAAAAGCCCTGCTTCTCAAGAGAGAGCTGACGTTGCG 6454
OY 981 tccacagcgcgtctcaaaccccgacactcttgccttaaaaagactgtagcaagtgac 1040
Db 6455 TCCACCAACCGCTCTCAACCCCTGCGACCTTCTGCTTAAGAGAGCTGATGAACCAAGTAC 6514
OY 1041 tcatgtatgccaatcaactatgattgtagagagactgtaggctccgcaagagaccttaacagat 1100
Db 6515 TCATATTTGCCATTCACATATTGATTTGAGAGAGACTGGGGTCCGCAAGAGACCTTACAGCAT 6574
OY 1101 accgctgactgagaaagagcttaacctgttactgacgagagagacgtatgtagtgaag 1160
Db 6575 AACCCTGACTGGAAGAGAGCTTAACCTGTTCTACTGACGAGAGAGAGATGTTGGTGAAG 6634
OY 1161 taagagagatgctgtaggagcagtgtagtgcgagagccgagacgtatgtagcagagcct 1220
Db 6635 TAAAGAGATGCTGGGGGCGGTGCTGAGACGGGACCCGACATCTGGGCCAGCAGCT 6694
OY 1221 gcccgaaggaacttcaagcgaagagcgtgactgactgagcccttaacgcaagcttgagct 1280
Db 6695 GCGCGAGAGAACTTCAGCAAAAGGCTGAGCTCATGCGCCCTACAGCAAGCTTTCGGCT 6754
OY 1281 gcccgaaggaagaaatcaaaacttatacagagcagtagtgccttgagactgca 1340
Db 6755 GCGCGAGAGAACTTCAGCAAAAGGCTGAGCTCATGCGCCCTACAGCAAGCTTTCGGCT 6814
OY 1341 cgtacagagggccatctataaacaaggggtgtgcttaacagagggagagaaataa 1400
Db 6815 CGTACATGGGGCATCTATAAACAAGGGGTTGCTTACCTAGAGAGAGAGAGAGATTA 6874
OY 1401 gaaacaagaggaatcttaagccttaagaaaccttaacattgccaanaagcttagctat 1460
Db 6875 GAACAAGAAGGAATTTCTAAGCTTATTGAACCCCTGATATTACCAAAAAGCTAGCTAT 6934
OY 1461 tatacactgctcctgagacatcaagaagcaagatctcatctcagaggaagcaagagagc 1500
Db 6935 TATACACTGCTCTGACATCAAGAAGCTAAAGATCTCTATCCAGAGAGAAACCAAGATGC 6994
OY 1521 taaccgggtgccaagcagagccagagctgtaacctctgaccttaataagaagc 1580
|||||

Db 6995 TGACCGGCTTGCCAAAGCAGGAGCCCAAGGCTTAACTTCTGCTATTAATAGAAATGCC 7054
OY 1581 caaagccccaagagaccagacagtagtacaccttaagagactggaagagataaagaat 1640
Db 7055 CAAGCCCGCAGAACCCAGACAGCATACCTTGAAGAGCTGGGCAAGATTAATAAAGAT 7114
OY 1641 agacagcttctctgagactccggagggagactgataactcataatgtagaagaaatcct 1700
Db 7115 AGACCA-TTCTCTAGACTCCGGAAGGAGACTGCTATACCTCAAGATGGAGAGAAATCT 7173
OY 1701 gccccacaagaagaggttagaataatgctcacac-agaataatgcttaacccactagaa 1759
Db 7174 GCGCCCAACAAAGAGGTTAGATATGTCACAAAGATATACCTGCTAACCCACTAGAA 7233
OY 1760 ctaaacacccgagcagcttgtagaagaacatcccttatcatgcttgcagagctacaagag 1819
Db 7234 CTAACACCTGACAGCATGCTGAGAAACATCCCTTATCATGCTTGAAGGCTACCAAGAG 7293
OY 1820 tggctgactggtggtcaaaacattgtgtgctctgacagctggttaatgcttaacctca 1879
Db 7294 TGCTGACTGCTGCTGCTGCTAAACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7353
OY 1880 gaatacctccaaggaagagactaagggagagcaccacagcgctactcggagagtgact 1939
Db 7354 GAATGCTCCAGGAGAGACTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7413
OY 1940 tcaactgagtagaagccgcttaataacggaacaaatataatggtttctgtagaacct 1999
Db 7414 TCACCTAGGTAAGGCGGCTGTAATATGCAAAATAATACCTATTGCTTTTGTACACACT 7473
OY 2000 ttcaagagtaggtagagcttactactaagaagaagacttcaacggtggtgctaga 2059
Db 7474 TTTCAAGATAGGTTAGAGGCTTATCTACTAAGAAAGAGACTTCAACCGTGTACTTAAA 7533
OY 2060 aaatcctgaggaagaaatlttccaagaattggaatacctaagagtaataggtgcaagatg 2119
Db 7534 AATATCTGAGAAATTTTTCAGATTTGGAATTAACCAAGATTAAGGTTGAGTCAAGCAT 7593
OY 2120 gtccagcttctgctcccaagtagtagtagagagagcggcgaagagattgaggatgtaga 2179
Db 7594 GTCCCTTTTGTGGCCAGGTAAGTCAAGAGACTGGCAGAGATTTGGGATTTGATTGGA 7653
OY 2180 aactgcatgtgcatacagaccccaagctcaagagcagtagagagatgaatagaacca 2239
Db 7654 AACTGCATTTGATACAGACCCCAAGCTCAGGACAGTAGAGAGATGAATAGAACCA 7713
OY 2240 ttaagaagaccccttaacaaatltgaaccaagaagactggaatlaatgattgtagctctcc 2299
Db 7714 TTTAAGAGACCTTACTTAATTAACCGCGGACTGGCTTATGATTTGATAGCTCTTC 7773
OY 2300 tgccttctgcttcttaaggttaggaagacaccccttgagacagtttggtcgaacccctatg 2359
Db 7774 TGCCCTTTTGTGCTTTTAAAGGTTAGGAACACCCCTGAGATTTGGCTGAGACCCCTATG 7833
OY 2360 aattgctctacggggagaccccccctgtgagcaaatgcttgccttgcacatagttgtagt 2419
Db 7834 AATTACTCTAGGGGAGACCCCCCTTGTGTAATTTGCTTCTATATAGTGTGTAAG 7893
OY 2420 tggcttctccagccttggcttcttaagcttaagagcgtcagatgggtgagagagag 2479
Db 7894 TGCTGCTTCCAGCTTTTGTCTCTAGGCTCAAGGCACTTGAGTGGGTGAGCAACAGAG 7953
OY 2480 cgtgagaagagctccggagagcctactc---agagagagacttgcaagttccacatgct 2536
Db 7954 CGTAGAGGAACTCCGGAGGCGCTACTAGAGAGAGAGAGACTTCAAGATCCCAATCCT 8013
OY 2537 tccaagttgagagatcaatataatgtagagcgaacggtgcaagaaac 2583
Db 8014 TCCAGTGGAGATTCAGTTACGTTAGACGCCACCTGACAGGAAC 8060
|||||

RESULT 3
US-08-766-528-2

Qy	2851	attactgagatcccgylacagylatlaatltaatltaacagcaaccacaaggagagctcccttggg	2910
Db	4925	actttactgactccgggtacaggtatTAATAATTAACAGCACTCAAGGGGAAGCTCTTGGG	4984
Qy	2911	gaacctggtgcctgaattatlatlgtctgccttcgaatcaaglaatccctgylctcatalgaca	2970
Db	4985	gacctggtgcctggaattATATATGTCTGCTTCATCATGAATACCTGGTCTCAATGACCA	5044
Qy	2971	ggccacaccccccgatgtactccgttcaagggttlaagtlgttgcacagaccaccaaa	3030
Db	5045	ggccacaccccccgatgtactccgttcaagggttlaagtlgttgcacagaccaccaaa	5104
Qy	3031	taattagaatatltgltgnaatccctcagaatttcccttgcagacaaatgagctgcataac	3090
Db	5105	TAATGAACAATATTGTGGAATCTCAGGATTTCTTTTGCAAGCAATGGACTGTGCATAAC	5164
Qy	3091	ttctaataatgtggaattgynaagggccagctctctcagaacagacagatgaattaccctt	3150
Db	5165	TTCTAATGATGGAATTTGGAAATGGGCACTCTCTCAGCAACACAGAGTAAAGTTACTTT	5224
Qy	3151	tggttaacaatccctaccagttatlaatcaatttaattatgccaatggagatgtgaagaatg	3210
Db	5225	TGTTAAACAATCCACAGTTATTAATCAATTAATTATGTGCGATGGGAGTGAAGATTGG	5284
Qy	3211	gcaacagcgggtacaaaagatgttcggaataaagcaataatagctgttcatcgttgaagct	3270
Db	5285	GCAACACCGGGTCAAAAAAGATGTGGAATAATCAACCAATTAAGCTGTCAATTCGTAGACT	5344
Qy	3271	agattccttaaaaaaataagtttctactaaaaaaaataaaaaaataaaaaa 3317	
Db	5345	AGATTACTTAAAAAATAGTTCCTACTGAAAAAAAGGAAAAACAAGAAAAA 5391	

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-075-272-1

Query Match	35.4%;	Score 1175.8;	DB 3;	Length 8655;
Best Local Similarity	66.1%;	Pred. No. 0;		
Matches 1778;	Conservative	0;	Mismatches 877.	Indels 36

[illegible]

Db 3995 ATGACTAATGACCGGATGACACACTATCAGAGTCTGCTGCTAAATGACCGGTAACCTTT 4054
 QY gttccaccagcgcctctcaaccctgcacactctctgcgtgaagaagactatgaaccagtg 1038
 Db 4055 GCGCCCTCTGCTCATTTTCAACCCAGCCACCTCTCTCTTAACGATGATTCGCTCCCA 4114
 QY 1039 actcatgattgcatcaactatgattgagagagctgggtccgcgaagacccttaagac 1098
 Db 4115 GTACATGATGTGACAGACATCTGCGGTGAGAAATTTGGGACCGACGAAAGACCTGACTGAC 4174
 QY 1099 ataccgctgaactgagaagtgtcaacctgttactgaagcaagcagctatgtgtgaa 1158
 Db 4175 CAACCTGCGCTG---AGCTCTAGTGTGTACACGATGACGATGACGATTTTCTGATTTGAG 4231
 QY 1159 gtaagaagatgctgtgggggagtggtgagcgggagccgcagctcgtgggcaagcagc 1218
 Db 4232 GGAAGAGGAGGCTGTGAGCTGCGGTGTGTGATGAGAAAAGTAAATTTGGGCAATGTCC 4291
 QY 1219 ctgcgggaagaaacttcaagcgaagaagctgaagcctcagcgaagcagcttgcg 1278
 Db 4292 TTGCTGTAAGAACTTTCGACAGAGGCTGAACCTCATAGCGCTTACACAGGCGCTCCGA 4351
 QY 1279 ctgagcgaaggaatccataaacttataagcagaagcagtgacttgcgactga 1338
 Db 4352 GAGGCCAAGGTAAAGATCAATTAATTTACACTGACAGCGCTATGCTTTGCTACTGCA 4411
 QY 1339 cagctacacgggcatctataaacaagggtgtgcttactcctcagcagggaggaata 1398
 Db 4412 CATATCATATGGGCGCATCTACAGGACGAGGCTTATTCAGCTTACACCGGTAAAGCATTC 4471
 QY 1399 aagaacaagaagaatcttcaagcctattagaagccttacttgcgaagaagcagct 1458
 Db 4472 AAAAACCAAGAAAGAAATTCCTGGCCCTGTWAGAGCCATACATGCGCCCAAGAGTAAGCC 4531
 QY 1459 attatacgtctctgtgaacttagaagaagccaaagatcctatatactagaaggaagcagtg 1518
 Db 4532 ATATCTCACTGTCTCCGGTCCAGAAAGAGAGGACTTGTGTGGCCCAAGGCAACCGGATG 4591
 QY 1519 gctgacgggttgcgaagcagcagcagcagcgttcaacctctgcctataatagaagc 1578
 Db 4592 GCAGACTCAGTACCAAAACAGGTGTGCTCAGGGGCGCATGATCTTAAGTAAAAAGAAAT 4651
 QY 1579 cccaaagcccaagaccagcagcagcagcagcagcagcagcagcagcagcagcagc 1638
 Db 4652 CC-----GTCCAAAGCCCTGA-----GGATGAAAGCATATGAT 4684
 QY 1639 atagaccagtctctgaagcctcggagggagcctctctataccctatctatgggaagaaatc 1698
 Db 4685 ATAAAGAACTATTTGGACTAGTATCCCTCCATCTTTTTCGAAAGGAAATATGAC 4744
 QY 1699 ctgcccacaagaaggttagaatacttcccaagacagatagcgttaaccacactaga 1758
 Db 4745 TTGACACCCGAGGAGGAATTAATTTTGGAAGAGACTACACAGTTTACCCACCTGGGA 4804
 QY 1759 actaaacacctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1818
 Db 4805 GTTGAAGAAATGATGAGACTGATTAAGATCCCGTATCAAGTCCCTTAAGTTGAAGTCA 4864
 QY 1819 gtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1878
 Db 4865 GTGGCTCAAAAGATTTAACTCTCTCAAGCGTGTGCACTTCACTAATCAACTAAACCC 4924
 QY 1879 agaatactccagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1938
 Db 4925 TACAAAGAACTGGAAGAGACAGCGGAGAACCGTCTCGAGTGTATTTGGAGGTGAT 4984
 QY 1939 tttacgtgagtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1998
 Db 4985 TTTACTGAAGTTAAACCCGGAATGTATGTAATTAAGTATCTGTATTTAGTATTTAGACACC 5044
 QY 1999 tttcagagtgagtgagcttactactaagaagaagcagcagcagcagcagcagcagcagc 2058
 Db 5045 TTTTCCAGATGGGTGAAGCGTTCCCTCAAACTGAGAGCTGCCAGATTGTGGCCAG 5104

QY 2059 aaatactgaggaattttcccaagatttgaatactaaagtaataagtcagacacat 2118
 Db 5105 AAGATTTTGAAGAAATCTCTGCCAAGATATGAGTACCTTAAGTAAATGGGGTCCCAAT 5164
 QY 2119 gtlccagcttctgttcccaagtaagtcaggaagcagcagcagcagcagcagcagcagc 2178
 Db 5165 GGACCAAGCTTTGTTTCCAGGTAAGTCAAGGATTTGGCCACTTATGCGCATTTATTTGG 5224
 QY 2179 aaactgacgttgcatacagaccacaagaagcagcagcagcagcagcagcagcagcagc 2238
 Db 5225 AAATTAATCTTGTCTTACCGCCCTCAAAAGCTCAGACAGTAAAGAGATGATGAGACA 5284
 QY 2239 attaaagagccttaccacaatttgcacacagagcagcagcagcagcagcagcagcagc 2298
 Db 5285 TTTAAAGGAACTTATTAATTTGSCATGTGAGACCGCGGGGAAAGACTGGGTGCTTC 5344
 QY 2299 ctgccccttctgtctttagagtgaggaacacccctgagcagcagcagcagcagcagcagc 2358
 Db 5345 CTCCCTCTTGGCTCTTCGAGACCGCGGAACACCCCGGACGTTTGGGCTCACTCTTTT 5404
 QY 2359 gaattgctctacggggagcccccgcgttgcagaatltg-----ccttgcacatagt 2412
 Db 5405 GAAATTCTGTATGGGGGACCTCCCTTTAAATAAAGATGTGGAACATTTGTTCCGAT 5464
 QY 2413 gctgaatgctgtcttcccaagccttctctcagcagcagcagcagcagcagcagcagc 2472
 Db 5465 TCAGGCTCTGTTCCT 5524
 QY 2473 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2532
 Db 5525 ACCCAAAATTTGGGACGACGATTAAGACGCGCTACACCCAGGAGCACCGGAGTACCCAC 5584
 QY 2533 cgttccaaagtgtgagatagcagcagcagcagcagcagcagcagcagcagcagcagc 2592
 Db 5585 GAGTTCAGAGTGGGGACCAAGTCTGTGACAGACATCAACCGGTAGGCTTGAACCA 5644
 QY 2593 cgttgaaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2652
 Db 5645 CGGTGGAAGGAGACCTATTATTTAGTGTACTAACACTCTCTACGGCAGTAAAGTGTACCGG 5704
 QY 2653 atcccacacctgatacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2703
 Db 5705 ATTGCTCTCTGATCAGCTTCAACAGCTCAAGAGGCGCCCGAGTCAAGAT 5755

RESULT 5
 US-09-111-085-3
 ; Sequence 3, Application US/09111085
 ; Patent No. 6100034
 ; GENERAL INFORMATION:
 ; APPLICANT: Stoye, Jonathan P
 ; APPLICANT: Weiss, Robin A
 ; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
 ; FILE REFERENCE: 4238/75168
 ; CURRENT APPLICATION NUMBER: US/09/111,085
 ; EARLIER APPLICATION NUMBER: 1998-07-07
 ; EARLIER FILING DATE: 1997-05-16
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 3482
 ; TYPE: DNA
 ; ORGANISM: Porcine retrovirus
 ; US-09-111-085-3

Query Match 34.1%; Score 1131; DB 3; Length 3482;
 Best Local Similarity 89.8%; Pred. No. 0;
 Matches 1251; Conservative 0; Mismatches 130; Indels 12; Gaps 3;

OY	1767	cccgacgaagcttggctcagaaacatccctcttaactaactcttcgaagcctacacgaagctgctcgga	18203
Db	6	ccctgcagcagcttgcctcagaacacccctcttaactctgtcttcgaaagctacacgaagctgcgcga	65
OY	1827	ctcgctgtgtccaaacatctgtctgcctctgcgcgcgcgtgtctaaatgtcttaactctccagaatacc	18866
Db	66	ctcgctgtgtccaaacatctgtctgcctctgcgcgcgcgtgtctaaatgtcttaactctccagaatacc	125
OY	1887	ttccaggaagaagagactaaaggggaagccacccagcgcgtccactctgggaagctggaattcaactga	19466
Db	126	ttccaggaagaagagactaaaggggaagccacccagcgcgtccactctgggaagctggaattcaactga	185
OY	1947	ggttaaaagccggccttaaaactaacgaaacaataatctatctgtgttttgttaagacactcttcag	20066
Db	186	ggttaaaagccggccttaaaactaacgaaacaataatctatctgtgttttgttaagacactcttcag	245
OY	2007	atgggttgaagagcttaactctctactaagaagaagactctcaacgcgtgtgtgcctaaagaatact	20666
Db	246	atgggttgaagagcttaactctctactaagaagaagactctcaacgcgtgtgtgcctaaagaataact	305
OY	2067	ggaagaagaattttcccaagattctggaatactcaagagtaaataggttcaagaacaaatggtccagc	21266
Db	306	ggaagaagaattttcccaagattctggaatactcaagagtaaataggttcaagaacaaatggtccagc	365
OY	2127	tttcgtctgcacaggtlaaggtcagagagacttgcgcacaaagataatttggagattgtatgtgaaactgc	21866
Db	366	tttcgtctgcacaggtlaaggtcagagagacttgcgcacaaagataatttggagattgtatgtgaaactgc	425
OY	2187	tttgcgtactaacaagaccccaaaagcctccagaacaagtttagagagagatagaatagaacattaaaga	22466
Db	426	tttgcgtactaacaagaccccaaaagcctccagaacaagtttagagagagatagaatagaacattaaaga	485
OY	2247	gaaccttacaacaatttgaccacaaagacatgcgaatlaaagtttgcgaatggtctctcccttcgcctt	23066
Db	486	gaaccttacaacaatttgaccacaaagacatgcgaatlaaagtttgcgaatggtctctcccttcgcctt	545
OY	2307	tgctgcttttagaggttagaggaacacccctctgacaagtttggcgtcgcgacccctatgtaattgtct	23666
Db	546	tgctgcttttagaggttagaggaacacccctctgacaagtttggcgtcgcgacccctatgtaattgtct	605
OY	2367	ctaaaggggagaccccccgcctctgcgaagaataatgcctcttgacacaaatgctgtaatgtctgct	24266
Db	606	ctaaaggggagaccccccgcctctgcgaagaataatgcctcttgacacaaatgctgtaatgtctgct	665
OY	2427	ttcccaagcctttgtctctctcctcagagctcaaaagcgtctgaagtggtgtgcagagcgagcgttgaa	24866
Db	666	ttcccaagcctttgtctctctcctcagagctcaaaagcgtctgaagtggtgtgcagagcgagcgttgaa	725
OY	2487	gcaacatccgggagagccctacacag--aagaaacttgcagaattccacaaatgccttccaaagt	25436
Db	726	gcaacatccgggagagccctacacag--aagaaacttgcagaattccacaaatgccttccaaagt	785
OY	2544	ttggaagatcagctctatagttagagcagccacgcgtgcagaagaaacctctcagactcggctggaag	26036
Db	786	ttggaagatcagctctatagttagagcagccacgcgtgcagaagaaacctctcagactcggctggaag	845
OY	2604	acctatctcgtacttcttgcacacacaaacgcgtctggaagaatcggaaggaatccccaactg	26636
Db	846	acctatctcgtacttcttgcacacacaaacgcgtctggaagaatcggaaggaatccccaactg	905
OY	2664	gattccatgcatacccaagcttaagagccgcgcacaccccgatctcggggttggaaagccggaanaa	27236
Db	906	gattccatgcatacccaagcttaagagccgcgcacaccccgatctcggggttggaaagccggaanaa	965
OY	2724	gaactgaagaatccccaataagtttgcgtccctccatactggtgttcccttaactctgtccaataactc	27836
Db	966	gaactgaagaatccccaataagtttgcgtccctccatactggtgttcccttaactctgtccaataactc	1025
OY	2784	ctcgaactaaatgtttaaagcgcctgtgtgcagacgcgcgaactccacataaaccttatctctca	28436
Db	1026	cccgagccaggtagtaaagccttatagacaagcttcgaacccccaatgaccttatccctta	1085
OY	2844	ccctggttacttactgactcgcgtgtacaggtatlaataatlaacagaacactcaaggggagagctc	29036

Accession	Sequence	Position
Db 1086	ccctgacgatattatgaccctgatagggtgtacctgtaataagaactcgaggtgtgtctc	1145
QY 2904	ccctgggaccctgtgtgacctgaattatattgtctgccttcgatcaagtaalccctgtgtca	2963
Db 1146	ctgaaggacacctgtgtgcctgaactcatcttcctgcctcgattgataaaccgcgtgtta	1205
QY 2964	atgacacggcgcaaccccccgatgcctacgtcgtctaaagggttttaagttgtggccagac	3022
Db 1206	-----aagacaaccccccaacctagctcgtatgtaatgaggttcattgtgcgccagg--	1251
QY 3024	cccccaataa tgaagaattatgtgtgaataacctcaagattctcttgcgaagcatgtgact	3083
Db 1258	-caacaggaagaaggaaatactgtgtggggtcttcggggaaatcctctctgtgagatgagct	1316
QY 3084	gcataacttcaatgatgtgggaattgtgaaatgtgccagctctctcaagaacagatgaagtc	3143
Db 1317	ggctgacactccaagatgagtgagactgtgaaatgtgcgactctctccagaacccggttaaat	1376
QY 3144	acctcttgttaa	3156
Db 1377	tctcctctgtccaa	1389

```

RESULT      6
US-08-716-351A-1
: Sequence 1, Application US/08716351A
: Patent No. 6033905
:
: GENERAL INFORMATION:
:
: APPLICANT:
:
: TITLE OF INVENTION:  Gibbon Ape Leukemia Virus-Based
:
: TITLE OF INVENTION:  Retroviral Vectors
:
: NUMBER OF SEQUENCES:  5
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE:  Floppy disk
:
: COMPUTER:  IBM PC compatible
:
: OPERATING SYSTEM:  PC-DOS/MS-DOS
:
: SOFTWARE:  PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER:  US/08/716,351A
:
: FILING DATE:
:
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER:  PCT/US94/03784
:
: FILING DATE:  06-APR-1994
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME:  Bastian,, Kevin L.
:
: REGISTRATION NUMBER:  34,774
:
: REFERENCE/DOCKET NUMBER:  15280-128-1PC
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE:  (415) 543-9600
:
: TELEFAX:  (415) 543-5043
:
: INFORMATION FOR SEQ ID NO:  1:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH:  8535 base pairs
:
: TYPE:  nucleic acid
:
: STRANDEDNESS:  single
:
: TOPOLOGY:  linear
:
: MOLECULE TYPE:  DNA (genomic)
:
: HYPOTHEICAL:  NO
:
: FEATURE:
:
: NAME/KEY:  misc.feature
:
: LOCATION:  1..8535
:
: OTHER INFORMATION:  /standard_name= "GalV SFAFO Genome"
:
US-08-716-351A-1

```

Db 3371 GATGCTTTTCTGCTCAGGCTACATCCCAACAGCCGCTGTTGCGTTTCGATG 3430
QY 82 agagatcagtgacggaagacccgagctcaccttgaccgagctgcccgaaggtc 141
Db 3431 AAGAGCCCAAGAAAAGGTACACAGCTGACCTGAGCCGCTACCAACAAAGGTTG 3490
QY 142 aagaccccccacacatcttgaagaaagctacacaggaaccttgacacactgagac 201
Db 3491 AAGAACTCTCCACTCTCTTCGACAGAGGCCCTCACCGAGATTTGGCTCCCTTTAGGCC 3550
QY 202 caaacacctcaagtgacacctctcctcagtaagtgatgacgtctctcgtgggagcaac 261
Db 3551 CTCAACCCCAAGTGTGTACTTCAATATGTGACGACCTCTTGCTGGCCGCCCA 3610
QY 262 aacacgagactgttagaagtgacgaagcactgctggaatgtctgacctgagac 321
Db 3611 TATGAAGACTGCAAAAAGAAACAGAAAGCTTTACAGAGATTATAGTATGGGATAC 3670
QY 322 aagacctgtcgaagaagcccaagattgcaagagaagtgaaacttgaggtacagt 381
Db 3671 CGGGTATCGGCTAAGAGAGCCAGCTGCGCAGAGAGAAAGTACCTTCTGGGGTACTA 3730
QY 382 ttgagggcgagcagcagtgctgtgacgaagcaggaagaaactgtatgctcagatagc 441
Db 3731 CTCAAGGAGAGAAAAGATGGCTAACCCAGCCGAAAGGCTACTGTATGAAAAATCCCT 3790
QY 442 gcccaacacacagccaacaaagtgagagattttgggagacgtgtgatttgagagctg 501
Db 3791 GTTCTACGAGCCCAAGAGAGCTCGTGAATTTCTAGGACATGCCGGATTCTGCAAGGCTC 3850
QY 502 tggatcccgagggttggaaccttagagcccaactctacccgctaaacaaagaaagag 561
Db 3851 TGGATCCCTGGGTTTGTCTTCCCTGCTGACCCCTGTGACCCCTTAACAAAGAGAGATC 3910
QY 562 ggaattcctcgtgagcagcaggaagcaggaagcatttgatgatacaaaagagccctg 621
Db 3911 CCTTTATTTGACTGAGAGAAATDAGCAGGCTTTTGACCAATTAATAAAAGCCCTTGCTG 3970
QY 622 agcgacactgtctgagccctccctcgtacgtacaaactaacaccttaccctttagtgatgag 681
Db 3971 TCAGGCCCTGCATTGGCCCTCCAGACCTCACMAAGCATTTACTCTATATATATAGATGAG 4030
QY 682 cgtaaaggagtagcccgagaggtttaacccaacccctaggaacataggaagagacgtgt 741
Db 4031 AGAGCCGGCGTGGCCCGGAGAGTGTCTACTCAGACTTTAGAGACCTGGCGGCGCCAGTA 4090
QY 742 gcttaacctgtcaaaagagctgtatcctgtacgagtggttggccgatagtctgaagct 801
Db 4091 GCATATCTATCAAAAAAACTGGATCCGGTGGCCAGCGGGTGGCCAACTGCTGTAAAGCG 4150
QY 802 atcgacgtgtgacatctgttcaagaagctgacaatgtgactttggagacagaata 861
Db 4151 GTTGCAGAGTAGACACTCTTCTCAAGSAGCCTATATAGTTAACTTGGGACAAATGTG 4210
QY 862 acgttaatagcccccacatgtatgagagaacatgltcggagcccccagacagatgag 921
Db 4211 ACTGTGATGTCTTCCATAGCCTCGAAAGCATCTGCGGCAACCCCGACCGGTGATG 4270
QY 922 accaagcccgacatgacacatcatcaagcctgtcttccagagagaggtgcacttgcct 981
Db 4271 ACCAATGCGCAATATGACTATTAACCAAGAGCTGCTGTAAATGAAGGATATGCTTTGGC 4330
QY 982 ccacacagccgctcacaacctcgaactctctgctgaagagagtgatgaacagagtagt 1041
Db 4331 CCCCCGTGTGCTTAACCAACGAGTACCTACTTCCATGCTGAGAGAGCCACCCAGTG 4390
QY 1042 catgattgcaatcaatattgattgagagacgtgggtccgcaagagccttaagacata 1101
Db 4391 CACAGGGTCTCAGAAATCTCTGCGGAGAAACTGGAACCTGACGAGACCTAGAGAACCA 4450
QY 1102 ccgctgtcgtgagagagtgatcgtgttcaacagcaggaagcagctatgtgtggagagct 1161
Db 4451 CCAATTGCCCGG--GGTGCAACCTGGTATACAGACGCTAGCAGTTTCATCAGCGAAGGT 4507

QY 1162 aagagatgctgggagcagtgatgtgacgggaccccgacagatctggccagcagctg 1221
Db 4508 AAGGAGAGAGAGGAGGCCCGGATGCTATGATGGCAAGCGGAGGTATGAGCTTACAGCTG 4567
QY 1222 ccggaaggaactcagcgaacaaagctgagctcaatgacccctcagcagcagcttggcgtg 1281
Db 4568 CCAGAGGTAGCTAGCGCCAGAAAGCTGACCTAGTACCTTTGACGAGGATTAAGCTG 4627
QY 1282 gcggaaggaatccataaacttatacgggaagcagtgatgctcttgctgacgacac 1341
Db 4628 GCCGAGGAAAAAATCAATCAATCTACACGAGCAGAGATGCTTTTGCCTGCTCAT 4687
QY 1342 gtacaggggacatctatacaaaaggggtgttatacctcagcaggaaggaagaaag 1401
Db 4688 ATTTATGGGCAATATATTAAGCAGAGGGGCTGCTCTCTCTGAAAAATATCAAA 4747
QY 1402 aacaaagagaatctcctaagccttataagagccttaccatttgccaaaagagctagctat 1461
Db 4748 AACAAAGAGGAATTTTGGCCCTGCTAGAGGCCATCATCTCCCTAGGGCGGTGCCATT 4807
QY 1462 atacactgtcctggacatacgaagaagcccaagatcatalctagagggaacagatgct 1521
Db 4808 ATCCACTGCTCTGGCCACAGAGGGGAAGTAACTGTGTGCTGGAACCGAGAGGCC 4867
QY 1522 gaccgggtgccaagcagcagccagcgtgttaaccttgctctataagaaagagcc 1581
Db 4868 GACGAGGCTCAAAAGCAGAGCCGCCCTGTCTCCACAGAGTGTGGCAGAA-----CTAC 4920
QY 1582 aaagcccccagaccagacagtagtacaccccttagaagacttgcaagagataaagata 1641
Db 4921 AAAACCTCAAGAGCAATTCAGCCCGCTCAAGAAAGACAGGCCGAGGGAG----- 4972
QY 1642 gaccagttccttgagactccggaaggagactgtatacctcatalatggaagaaatcgt 1701
Db 4973 -----CTC 4975
QY 1702 ccccaagaagaaggttagaatactgccaacagatacatcgtctaaacacactaggaact 1761
Db 4976 ACCCTGACCGGGGAAAGATTCATTAAGCGTTACATCAGTTAACTCACTTAGAGCA 5035
QY 1762 aaacacctgacagatgtgtcagaacatccctctacatgttcttgaggttaccagagtg 1821
Db 5036 GAAAGGCTTCTCCACTAGTAGAACCGACCAAGCTCTCTCAACCACTCCAAATGTGA 5095
QY 1822 gctgactgggtgtcaaacatgtgtgcccctgcagcgtgttaatgtactaactccaga 1881
Db 5096 GTTCGGGAAGTCAACAGTCAAGTGTGAGGCTTGTGCATGACTAATGCGGTCAACCACTAC 5155
QY 1882 ataccctcaggaagagactaaagggaagccaccccgagctccactgggaagtgatc 1941
Db 5156 AGAGACACCGGAAAGAGCAACAGAGATGCAACCGCGTGTACGGGAGGATGACTTTC 5215
QY 1942 actgaagtaagccgagctaaatatacgaacaaatactatgtgttcttgtagacactt 2001
Db 5216 ACAGAAATTAAGCGCTGTGTGTATGGAACAAAGTATCTTTAGTATCTATGATACTTTC 5275
QY 2002 tcaagatgggttagaggttatacctactaagaaagagacttcaacgtgggtgtaga 2061
Db 5276 TCCGATGTGGTAAAGCTTTTCTTCAAAACAGAAACGCGCTAATGCTGTGTAATAA 5335
QY 2062 atactgaggaagatttccaaagatttgaaacttaaggttaaggtcaaacatgt 2121
Db 5336 ATATTGAGAAAGAAATTTTACCCTGCTTCCGGATCCCTTAAGGTCTGGGTCAAGATGGC 5395
QY 2122 ccagcttctgtgcccaggttaagtcaggaacttgccaaagatactggagattgattgaa 2181
Db 5396 CCGGCTTGTGTGCTAGGATTAAGTCAAGGACGTGGCACATCACTGCGGGATTAATTGGAAG 5455
QY 2182 ctgcatctgtgatacagaccccaagctcaggaacaggttagagagagataaataagcact 2241
Db 5456 TTACATTGTGCTATAGACCCCAAGCTCAGGTGAGTGAAGAAAGATGACAGAACAATTT 5515

Db 3889 CTAGTCATCTGTGCCCCCATCATGAGCAGTGTAAAGCAACCCCTGATGCTGG 3948
Qy 919 atgaccagaagcccgatgaccacatacaaacctgtctc---caagagaaggtact 975
Db 3949 CTCTCCAGGCCGATGACCCATGACAGGCTCTGCTTGTGACGAGACGATCGATCGAG 4008
Qy 976 ttgcgtccacacacgctctcaacctgcacctcttcctgtaagagataatgaaca 1035
Db 4009 TTGGAGCCCAATAGTGGCCCTTAACCCAGCTAGCTGCTCTCTTA---CCTAGAGAGGG 4065
Qy 1036 gtgactcatgactgcatcaactatgtatgagagagactgggtccgaagaacctaca 1095
Db 4066 CTGGCAACATGACTGCTTGAATCTTGCTGAGGCCACGAACTAAGACCAATCTTACG 4125
Qy 1096 gacatacgcgtgactgagagaagtgttaacctgttcaactgacgagaagcactatgtgt 1155
Db 4126 GACCAACCTCTCCAGACCTGACCAACACCTGTACACAGATGGAGAGCTTCTTGCA 4185
Qy 1156 gaagtaagagatggtgtggcgcaagtgtgtgaogggagaccgacgactgtggccagc 1215
Db 4186 GAGGGGACGCGACAGGCGGAGCAGTAAACCAACCGAGACCGAGGTAGTGTGGCCAAA 4245
Qy 1216 agcgtccggaagaaactcaagcgaagagctgactgctcctcagcaagcttgy 1275
Db 4246 GCACCTCCAGCCGGGACATCGGCCCAAGAGCTGAGTTATAGCGCTCACCAACCTTA 4305
Qy 1276 cggctggcgaaggaatcaataacattatacagagacgaagtatgcttggagct 1335
Db 4306 AAATATGCGAAGGTAAAGAGCTGAATGTTTACCGAGATGCCGTTATCTTTCCTTCC 4365
Qy 1336 gcaacagtacaagggccatctatacaaaaggggtgttgaactcgaagagggaa 1395
Db 4366 GCCCATATTCACGGAGAAATATATGAAAGCGCGGCTTCATCATCAAAAGAAAGAA 4425
Qy 1396 ataaagaacaaaggaatcttaagcctattagaagccttacttggccaagagcta 1455
Db 4426 ATCAAAATATAGGACGAGATCTTGGCCCTACTGMAAGCTCTTCTGCCCCAAAAGACTT 4485
Qy 1456 gctattatacactgtctcgtgacatacagaagcgaagatcatalctatagaggaagcag 1515
Db 4486 AGCATATATTCATTTGCCGGGACATCAGAAAGCGGAGGCAAGGAGGCAAG 4545
Qy 1516 atggctgacccgggttgcgaagcagcccgctgttaacctctgcctataagaa 1575
Db 4546 ATGGCGACCAAGCGGCCGAGAGTAGCCACTAGAGAAATCCAGAGCTTCCACACTT 4605
Qy 1576 acgccaagaagccagaagccc--agaagacagtaacacctagaagactgcaagagataaa 1634
Db 4606 CTGATAGAAATTCAGCCCTTATCTACTGTAACATTTTCACTATACGCTGACTGACATA 4665
Qy 1635 aaagatagaccagttctctgagactc-----cggagggagacgtgctatacctcatat 1686
Db 4666 AAGATCTGACTAATACTAGGGGCACTTATGAGATGCAAGAGAGTGTGGTTATACG 4725
Qy 1687 gggagaagaatcctgcgcccaagaaggggttaagaatagtccaacagataatcgtcta 1746
Db 4726 GGAAGACCTGTATAGCTCATATTCACCTTTGAACTATTAATTTTTCATCATCAATGG 4785
Qy 1747 accaccctag-----aactaaacacgtgcagcaggtgtgtcagaacatcccttatcat 1800
Db 4786 ACCACCTCAGTTTCTCAAAAACAAAGGCTCTTTAGAAAGAGACTACTGCTCTTATTTAC 4845
Qy 1801 gtctgaaggtacacagagtggtgactcgtgtgtaaacattgtgtccctgcagctg 1860
Db 4846 ATGCTGAACCGGATCGAAGCGTCAAGACATCAGACTGAGACTTGCCAAAGCTGTGACAG 4905
Qy 1861 gtaatgtcaatccttcagaatacctccaagaaagagactaaggggaagccacccagc 1920
Db 4906 GTCAATGCGACGAAGTCTCCGCTCAAAACAAAGGACTAGAGTTCTGAGGGGACGACCCGCG 4965
Qy 1921 gtcactgggaagtggaactcaactgaggtgaagcgggtaaatagaaacaaatatac 1980

Db 4966 ACCCACTGGGAAATTTGATTTTCACTAGAGTAACCTGGCTGTATGGGTATTAATATCTT 5025
Qy 1981 ttggtttttgtagacaccttttcaggatgggtaagagcttaactactaaagaagagact 2040
Db 5026 TTAGTTTTCATAGACACTTTTCTTGATGGGTATAGACTTTTCCAAACCAAGAAAGAACT 5085
Qy 2041 tcaacgtgtgtgtagaataatcagagaaatcttccaaagatttgaaatcctaag 2100
Db 5086 GCCAAAGTTTGAACCAAGAACTACTAGAAAGAAATTTTCCCAAGATTTCGGCATCCACAG 5145
Qy 2101 gtaataggtaagaacaatgtgtccagcttctgttgcgaagtaagtaagagagactggc 2160
Db 5146 GTATTGGGAAACCGCAATAGGCGCTGCTGCTTCCAAAGTAAGTAAGAGATGAGCCGAT 5205
Qy 2161 atattgggagattgtagaactgtagatgtgtacagagaccgaagcctaagagagta 2220
Db 5206 TTACTGGGGGTTGATTTGAAACTACATTTGCTTACAGACCCGAGATTCAAGGCAAGTA 5265
Qy 2221 gagaagatgaatagaacatlaagagaccccttaccatgaacacagagactgcat 2280
Db 5266 GAAGAATGATATAGGACATCAAGAGACTTTAACTAATTTGAGCTTGGCACTGGCTCT 5325
Qy 2281 aatgattgtagtgcctcctgccccttgtgtcttttaggtgaggaacaccttgagcag 2340
Db 5326 AGGCACTGGTGTCTCTGCTTCCCTAGCCCTGATGAGGCCGCAACACGCGGCGCC 5385
Qy 2341 ttgggtgacccctatgaatgtccttaaggggagaccccccgttggcagaatgtgc 2400
Db 5386 CATGGTCAACCCCATAGAAATCTTATAGGGGACCCCGCCCTTGTAACCTTCCCT 5445
Qy 2401 ttgcacatagtgtagtgccttccacagccttgttcttagtgcacagcgtc 2460
Db 5446 GATCTGACATGCGAAGAGTTACTCATTAACCCCTCTCCAGGCCCATTTAGAGGCACTC 5505
Qy 2461 gattgtgtgagcagcagcgttggaagcagctccgggagagcctactcaagagagagactg 2520
Db 5506 TACCTGCTCCAGCAAGAGTCTGAGACCCGTTGGCGGAGCTTCCAAAGCAACTGAC 5565
Qy 2521 c-----aagttcacatgcctccaagtgtgagatcagctatgtttagagcgcacgct 2574
Db 5566 CGCGCGTAGAGCTCTCAACCTTCCGAGTGGTACACAGTGGGTCCGACACACAA 5625
Qy 2575 gcaggaacacgtcagagactcgttggaagagacattatcgtacttggcacacacag 2634
Db 5626 ACTAATAATCTTAGAACCCTGGAAGAGACTTATACCGTCTACTGACTACCCCAAC 5685
Qy 2635 gctgtgaagtcgaagaalcccaactgtaaccatgcatcccaagcttaagcggc--- 2690
Db 5686 GCTCTCAAGAGTGAAGGCAATTTGACGCGTGAATCCAGCGTGCCCAAGTAAGGCTGCCGAC 5745
Qy 2691 -----gcacatcccgatcggggttgaagagcgaagaaagactgaataccctc 2739
Db 5746 ACCAGATTTGAGCCACCTCGGAATCGACATGAGGCTGTCAACCTCTCAAAATCCCTTA 5805
Qy 2740 aagcttgcctcatcggtgtgttcttactactgttcaataactcctcaagttatgttaa 2799
Db 5806 AAGATAAATTGACCCGCGGAGACTCTTAATCCCTTAATTTCTTCTGTCTGCTCAAG 5865
Qy 2800 aagccttggagagccc 2817
Db 5866 GGGCCAGATCCGCAAGC 5883

RESULT 8
US-08-886-642-8
: Sequence 8, Application US/08886642
: Patent No. 5952474
: GENERAL INFORMATION:
: APPLICANT: Pinter, Abraham
: APPLICANT: Kayman, Samuel
: TITLE OF INVENTION: FUSION GLYCOPROTEINS
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.
STREET: 45 Rockefeller Plaza, Suite 2800
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/886,642
FILING DATE: 01-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/110,300
FILING DATE: 20-AUG-1993
APPLICATION NUMBER: 07/938,100
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hone, William J.
REGISTRATION NUMBER: 26,739
REFERENCE/DOCKET NUMBER: 0763/010002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212/765-5070
TELEFAX: 212/258-2291
INFORMATION FOR SRO ID NO: 8
SEQUENCE CHARACTERISTICS:
LENGTH: 8323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-886-642-8

Query Match 28.1%; Score 933.6; DB 2; Length 8323;
Best Local Similarity 60.0%; Pred. No. 1.2e-273;
Matches 1702; Conservative 0; Mismatches 1094; Indels 42; Gaps 7;

QY 19 acagatgccttctctcgtcgtgagattacacccacatagccacacatttctgctcga 78
DB 3049 AAAAATCTTTTCTGCTGAGACTCCACCACCCAGTCAGTCTCTTCCGCTTTGAG 3108
QY 79 tggagatccaggtacgggaagaacccgggacgtcaccctcgagcccgactgcccaga 138
DB 3109 TGGGAGATCCAGAGATGGGAATCTCAGGACAAATTAACCTGAGACACCTCCGCGAGGT 3168
QY 139 ttcaagaactcccgacacatctttgacgaagccctacacagagacctggcacaattcag 198
DB 3169 TTCAAAAACAGTCCACACCTCTTTGATGAAAGCCCTGCACAGGACCTCGCAGACTTCGG 3228
QY 199 atccaacacctagtgacccctccctccagtagcgtgatgactgcttcgcgcgagcc 258
DB 3229 ATCCAGACCCAGACCTGATTTGCTCCAGTATGTAGTACTTACTGCTGGCGCCACT 3288
QY 259 accaacaagactgcttagaaggtacgaagcaactcgtggaattgttcgaactlag 318
DB 3289 TCTGAGCTTACGTCAACAGGTACGCGGCGCTGTTACAAACCTTAGGGGACCTCGGA 3348
QY 319 taagaagcctctgtaagaagccagatttgaagaagaggttaacataacttgggtac 378
DB 3349 TATCGGCTCTGGCCAGAAAGCCCAAAATTTGCCAGAAACAGGTCAAGTCTTGGGGTAT 3408
QY 379 agtttcggggcgagcagcatgctgaacgagcagcgaagaacactgagtcagata 438
DB 3409 CTCTTAAAGAGGGTCAAGATGCTGATGAGCCAGAAAGAGACTGTGATGGGGCGAG 3468
QY 439 cgggcccacacagccacaagaatgagagagtttttgggacagctgatttgcaga 498
DB 3469 CCTACTCCGAAGACCCCTCGACAACTAAGGAGTCTCTAGGAGCGGAGGCTTCTGTCC 3528
QY 499 ctgtgatcccggtgttgacacttagacgcccacactacccgctaaccaagaanaa 558

DB 3529 CTCTGATCCCTGGGTTTGAGAAATGCGAGCCCTTCTTACCTCTACCAAAAGCGGG 3588
QY 559 gggggaattctcctgggctccctgagcaacgaagactttagtactaacaagccctg 618
DB 3589 ACTCTGTTGAGTGGGGGCCAGACAGCAAAAGCCCTACCAGAGATCAAGAGCTCTC 3648
QY 619 ctgagcagactgctcgtgcccctcctgaagtaactaaaccttaaccttaattgagat 678
DB 3649 TTAAGTCCCTGCTCCCTGGGATTTGACAGTTCAGAACCTTCTTGTGTGAC 3708
QY 679 gagcgaagggagtagccggagaggtttaaccacaaccttgaccatggggagagact 738
DB 3709 GAGAGACGAGGCTACGCCCAAGGTGCTTACGCAAAACCTGGGCTTGGCTGGCGCG 3768
QY 739 gttgcttaactgtcaagaagacttgaacctgaagcagtggtttggccgtatgctgaag 798
DB 3769 GTGGCTACTCTGTCACAAAAGCTAGACCCAGTGGAGCTGGGTGGCCCTTGCTACGG 3828
QY 799 gctatgcagctgtgacatactgtlcaagaagcgtgacaaattgacttggagacgaat 858
DB 3829 ATGATAGCAGCCATTCGCGCTTGACCAAAAGAGCTGGCAAGCTACCATGGAGACGCA 3888
QY 859 ataactgtaatgccccccatgcatgtgggaacatcgttcgagcagccccaagcagatg 918
DB 3889 CTAGTATTCTGGCCCCCATGCGATGAGGCACTAGTTAAACCAACCCCTGATGCTGG 3948
QY 919 atgaccaagccgcagatgacccacatcaaaagcgtgctct--cacagaagggtaact 975
DB 3949 CTCTCCAGCGCCGATGACCCAGTACACAGGCTGTCTCTTGACACGAGCAGCATTCAG 4008
QY 976 ttgcgtcaacagcgcgtctcaacctgcaactcttcgtcgtgaagagactlgaagaca 1035
DB 4009 TTGGAACCAATGATGGCCCTTAACCCAGCTACGCTGCTCCCTCTA--CTTAGAGAGGG 4065
QY 1036 gtgacctagatgtgccaataactattgttagggagactggttcgcaagacttaca 1095
DB 4066 CTGCAACATGACTGCTTTGACATCTTGCTGAAGCCACGGAACCTAGACACATCTTACG 4125
QY 1096 gacataccgcttacttgaagaagtgttaacctgttcaactgaaggaagcagctatgtg 1155
DB 4126 GACACACCTCTGCCAGAGCGTACCAACACCTGTGTACAGAGTGGAGAGCAGCTCTGCA 4185
QY 1156 gaaggtlaagaagatgcttggggcgagtggtgagcgggacccgacagatcgtggcagc 1215
DB 4186 GAGGGGACGCGCAAGCGGAGCAGACAGTAACCAACGAGCAGAGTGTGCTGGGCCAA 4245
QY 1216 agcctgcccgaaggaacttcagcgcaaaagcttagctcactgagccctcaagcagctt 1275
DB 4246 GCACCTGCCAGCCGGGACATCGGCCCAAAAGAGCTGATTGATAGCCCTCACCAAGCCCTTA 4305
QY 1276 cggctggccgaaggaataatccataaacttaacgaagcagatgacttggcttggagct 1335
DB 4306 AAAAATGGCAGAAAGTAAAGAGTGAATTTTACCCGATAGCCGTTATGCTTTGGCCACT 4365
QY 1336 gacacgtacacaggggcatctataaacaagaagggtgttcttacctcagcaggaagga 1395
DB 4366 GCCCATATTCAAGGAGAAATATATATGAAGGGCGGGTGTCTACATCATCGAAGGAAGAA 4425
QY 1396 ataaagaacaaaggaataatctlaagctatlaagaagccttaacttgcacaaaggcta 1455
DB 4426 ATCAAAAATTAAGAGAGATCTTGGCCCTTACTGAAGGCTCTCTTCTTCCCAAAAGACTT 4485
QY 1456 gctatataactgtccctgacatataagaagaacgaagttccatatatagaagggaacag 1515
DB 4486 AGCATTAATTCAATTGCCCCGGGACATCAGAGGAAACCGCGGAGGCAAGGGGCAACAGG 4545
QY 1516 atgctgaacgggttgcgaacagcagccagcagctgttaaccttccctacataatgaa 1575
DB 4546 ATGGCGGACCAAGGGGCCGAGAGTATGACCACTAGAGAAATCCAGAGACTTCCACACTT 4605
QY 1576 acgccaagaagcccaagagcc--agaagcagtlacaccttagaagacttgcgaagataaa 1634

Db 3707 CTAGTCAATCTGGCCCCCGCCAGTAGAGGCACTAGTTAAGCAACCCCTGATCTGCTGG 3648
Qy 919 atgaaccaagcccgatgagcccatcaagaagctctctc---caagaagaggtcact 975
Db 3647 CTCTCCCAAGCCCGAAGAGCCACTACAGGCTCTGCTTGTGACGAGCAACCCAGCCAG 3588
Qy 976 ttgcctcacagccgctcccaaccctccctctctgctgaagagactgtgaca 1035
Db 3587 TTGGGACCAATATGAGCCCTTAACACGCTACCTGCTCTCTTA---CTGAGGAGGG 3531
Qy 1036 gtactcaatgatgtgcaatcaactatgtatgaagagactgtggtccgcaagagcttaca 1095
Db 3530 CTGCAATGATGAGCTGCTTGTACATCTTGGCTGAAGCCGAGCACTAGACAGATCTTAGG 3471
Qy 1096 gacataccgctgactgtgagaagtgtcactgtgttactgtgaggaagcagctatgtgtgt 1155
Db 3470 GACCAAGCTCTCCAGAGCTGACACACCTGTACACAGATGGAGCAGCTTCTGCA 3411
Qy 1156 gaaggtaaagagatgt 1215
Db 3410 GAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3351
Qy 1216 agcctgtcggaaggaactcaagcgaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1275
Db 3350 GCACTGCGACGCGGAGCATGCGCCCAAGAGCTGATGTATGAGCTCACCCAGGCTTA 3291
Qy 1276 cagctgtcggaaggaatccataaacttataagcaagagagatgtgtgtgtgtgtgtgt 1335
Db 3290 AAATGCGAGAGGTAAAGAGTGAATGTATACCGATATGCGGTATGCTTTGCGACT 3231
Qy 1336 gcaacagtaacaggggcaatcaaaagaaggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1395
Db 3230 GCCCATATTCACGAGAGATATATAGAGGCGGGTGTGTACATCAGAGAGAGAGAGAA 3171
Qy 1396 ataaagaacaaggaagaaattcaagccttataagagccttaccattgtcccaaaagctta 1455
Db 3170 ATCAAAATTAAGGAGGAGTCTTGCCCTTACTGAAGGCTCTTCTTCCGCAAAAGACT 3111
Qy 1456 gctatitacactgtctcgtgacatcagaagaagcaagatctatctatctagaagacag 1515
Db 3110 AGCATTAATTAATGCTCCGCGGACATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3051
Qy 1516 atggtctgacgggt 1575
Db 3050 ATGGCCGACCAAGGCGCCGAGAGTAGCAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2991
Qy 1576 acgcccagaagcccaagaagcc-agaagacagtaacccctagaagactgtgcaagaagataa 1634
Db 2990 CTGATAGAAAATTCAGCCCTTACTACATGAACATTTTCACTATACGGTGAAGTGA 2931
Qy 1635 aaagatgaccaggtctctgtgagctc-----cggaagggaccgtgtatctatctat 1686
Db 2930 AAAGATCTGACTAATAGAGGCGGACCTTATGAGATGCAAGAGAGAGAGAGAGAGAGAG 2871
Qy 1687 ggggaaggaatctcgtcccaacaagaaggggttagaatagtgtccacaagataatcgtctta 1746
Db 2870 GGAAGGCTGTAAATGCTGATCAATTCACCTTGAACCTATTAGATTTTCTTCAATCA 2811
Qy 1747 acccaactag-----aactaaacctgtgcaagactgtgtcagaacatccctctatct 1800
Db 2810 ACCCACTCAAGTTTCTCAAAAACAAAGGCTTCTTAGAAAGAACTACTGCTTATATAC 2751
Qy 1801 gtctgtgagctaacaggaagt 1860
Db 2750 ATGCTGAACCGGATGCAAGGCTCAAGAGATCACTGAGACTTGCACAGCTGTGACAG 2691
Qy 1861 gtaatgtcaatctctccacaataactcctcagaagaagacttaaggggaagcccaagc 1920
Db 2690 GTCAATGCCAGCAAGTCTGCGCTCAAAACAGGAGCTAGAGTTGAGAGGACCCAGCCGCG 2631
Qy 1921 gctacatgggaagtgtgacttactgtgaggtgaagccggttaatacaggaacaatatctta 1980

Db 2630 ACCCACTGGAAATGATTTCACTGAGGTAAACCTGGCTGTATAGGTATATATCTT 2571
Qy 1981 ttggtttttagaacactcttcaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2040
Db 2570 TTAGTTTTCATGACACTTTCTGTGATGGGTAGAGCTTTTCCACCAACAAAGAAACT 2511
Qy 2041 tcaacctgt 2100
Db 2510 GCCAAAGTTGTAAACCAAGAGCTACTAGAGAAATCTTCCCAATTCGGCATGCCACAG 2451
Qy 2101 gtaataaggtcagaacatgt 2160
Db 2450 GTATTGGGAGACCGCAAGAGGCGCTTCCCTTCCCAAGGTAGTCAACAGTATGAGCGAT 2391
Qy 2161 atattgggattgt 2220
Db 2390 TTATTGGGGGTTGATTGGAATCTACTATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2331
Qy 2221 gaaagatgaatagaaacattaaagaagcccttaacaaatltgacaaagaagactgtgcaat 2280
Db 2330 GAAAGATGAATATAGACATCAAGAGACTTAACTAATTAAGAGCTTGAACCTGCTCT 2271
Qy 2281 aatgtatgt 2340
Db 2270 AAGGAGCTGGGTCT 2211
Qy 2341 ttgt 2400
Db 2210 CAGGCTTCACCCCATATGAAATCTTATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2151
Qy 2401 ttgt 2460
Db 2150 GATCTGTACAGAGGAGAGAGTACTATACATCCCTCTCTCAAGCTCATTTACAGGACTC 2091
Qy 2461 gagt 2514
Db 2090 TACCTGTACAGAGAGAGAGT 2031
Qy 2515 gactgtgcaaggtctccatcgtctcgaagtgtgagatltcagctatgtgtgtgtgtgtgtgt 2574
Db 2030 CGGCGGATAGTACCTACACCTTCCGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1971
Qy 2575 gcaagaacactgtgagactcgt 2634
Db 1970 ACTAAATTAATGAAG 1911
Qy 2635 gctgtgaaggt 2690
Db 1910 GCTCTGAAGTGAAG 1851
Qy 2691 -----gcacactccgattcgt 2739
Db 1850 ACCAGAGATGAGCCACAGAGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
Qy 2740 aagctgt 2799
Db 1790 AAGATTAAGTGAAG 1731
Qy 2800 acgctgt 2817
Db 1730 GGGCCAGATCCGACAGC 1713

RESULT 11

US-08-886-642-9/c
; Sequence 9, Application US/08886642

; Patent No. 5952474

; GENERAL INFORMATION:

; APPLICANT: Pinter, Abraham

; APPLICANT: Kayman, Samuel

; TITLE OF INVENTION: FUSION GLYCOPROTEINS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:


```

Db 2990 CTGATAGAAAAATTCAGCCCCCTTACTATGACATTTTCTACTATAGCGTACTGACATA 2931
QY 1635 aaagatagaccagttctctgagctc-----cgaggggacctctatactcatat 1666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2930 AAAGATCTGACTTAACATAGGGGCCACTTATACACATGCAAGAGTGTGGTTTATCAG 2871
QY 1667 gggagaagaatccctgcccacaaagaggttagaataatgtaacagatatccagtccta 1746
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2870 GCAAAAGCCTGTAATGCCCTGATCATATACCTTTGACATTTAGATTTTCTTCATCATATTG 2811
QY 1747 acccaacctagg-----aactaaacacctgacagcagttgtgcagaacatccctatcat 1800
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2810 ACCCAACCTCACTTTCTCTCAAAAAAAGAGCTCTTCTAGAAAAGAACTACTGCTCTTATTAC 2751
QY 1801 gtctgaggtctaccagagagtgctgactcggtgtgtcaaacattgtgaccttcgagctg 1660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2750 ATGCTGAACCGGATTCAGACGCTCAAGACATCTAGACTTGTCCAGCGCTGTGCACAG 2691
QY 1861 gtaattgtaatacctcctcagaatacctcagaagaagactaaaggagagccacccagc 1920
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2690 GTCAATGCCAGCAAGTCTGCCGTCAACAAAGGAGCTAGAGTTGAGGGGACCCGACCGGCG 2631
QY 1921 gctacatgggaagtggagcttcacagtaagtaagccgctaaatacggaaacaatatcta 1860
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2630 ACCCACTGGGAATTTGATTTCTACTGAGGTAAACCTGGCCTGTATGGGTAAATATCTT 2571
QY 1981 ttggtttttttagacacatttccagataggtagaggtatctatcctactaaagaagagct 2040
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2570 TTAGTTTTCATAGACATTTCTCTCGATGGGTAGAAAGCTTCCCAACCAAGAAAGAACT 2511
QY 2041 tcaacctgtgtgctaaagaataactcggaggaatttttccaaagatttgaatacctaag 2100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2510 GCCAAATTTTAAACCAAGAGCTACTAGAAATATCTTCCCGAGATTCCGCGATGCCACAG 2451
QY 2101 gtaataagtcacacaaatgtgccagcttcgttgcacagtaagtaaggagcagcagc 2160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2450 GTATTGGGAACCGACATGGGCGCTGCTTGTCTCCAGGTAAGTAGACAGTAGAGCGCAT 2391
QY 2161 atattggagatctgtagaaactgcatctgtcatcacagcccaaaagctcagagcagcagta 2220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2390 TTATTGGGGCTTATGGAATCTATCATTTGCTTACAGACCCGAGAGTTCAAGGTCAAGTA 2331
QY 2221 gagaggtatagtaaaccttaagaagaccttaacaaattgacacagagagctgcat 2280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2330 GAAGAAATGATAGCAATCAAGAGACTTAACTTAACATAATGACGCTTCAACTGCTCT 2271
QY 2281 aatgatagatgctcctcctgaccttgttctttagaggtgagaaacccctttagag 2340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2270 AGGAGACTGGGTGCTCTGCTTCCCTTGGCCTTATCAGACCCGCAACAGCCGGGCGCC 2211
QY 2341 ttggagctgaacccctatgaattgctctacgggggagcccccgcgttggcagaaattgc 2400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2210 CAGGGTCTACCCCATATGAATCTTATATGGGGGACCCCGCCCTGTAAACTTCCCT 2151
QY 2401 ttggacataatgctgagtgtgctgcttccagccttgttctctagagctcaagcgctc 2460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2150 GATCCTGACATGGCAAAAGTTACTCTAAACCCCTCTCTCAAGCTCATTTACAGAGCAC 2091
QY 2461 gaagtggttagagcagcagcgtlbgaaagcagctccggagagcctactcagc-----agga 2514
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2090 TACCTGGTCCAGCAGACAGTCTGAGACCGTTGGCGGAGCTTACCAAGACGAACTAGAC 2031
QY 2515 gacttgcaagttccacatcgcttccaaagttgagattcagtcaltgttagcgcacagt 2574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2030 CGGCCGCTGATACCTTCTCCGGGTGGTGACACAGTGTGGGTCCGCGAGACCA 1971
QY 2575 gcaagaaacctcgaagactcgtgfggaagagccttactctgtactctttagccacacaa 2634
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1970 ACTAAAAATCTTAGAACCCCGCTGGAAAGAGACCTTATACCGTCTACTAGTACACCCCA 1911
QY 2635 gctgtgaaagtcgaagaaatcccaactgtagatcagtcacacagtttaagccgagc----- 2690
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1910 GCTCTCAAGTAGACGGCATTTGCACGCTGATCCACGCTGCCACGTAAAGGCTGCGCAG 1851

```

```

QY 2691 -----gccacctccagattcgggtlbgaaagccgaaagaactlgaataatccctt 2739
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1850 ACCGAGATTGAGCCACGACAGATGACATGCGCTGTTCACAGCGTCCCAAAATCCCTTA 1791
QY 2740 aagcttcgctccatcagcgtgtgttcttactctgtcaataactctcaagttatggtaa 2799
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1790 AATATAGATTGACCCCGGAGACCTCTAATCCCTTAATCTCTCTGCTGTCAAG 1731
QY 2800 acgccttlygacagcc 2817
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1730 GGCGCAGATCCGACGAC 1713

```

```

RESULT 12
PCT-US93-08041-9/c
; Sequence 9, Application PC/TUS9308041
; GENERAL INFORMATION:

```

```

; APPLICANT: The Public Health Research Institute of the City of
; APPLICANT: New York, Inc.
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull and Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08041
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,100
; FILING DATE: 28-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/-----
; FILING DATE: 20-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H.
; REGISTRATION NUMBER: 32,140
; REFERENCE/DOCKET NUMBER: 11698A50
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10367 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US93-08041-9

```

```

Query Match

```

```

Best Local Similarity 59.98; Score 932; DB 5; Length 10367;
Matches 1701; Conservative 0; Mismatches 1095; Indels 42; Gaps 7;

```

```

QY 19 acagatgcctctctcgtcgtgagattacaccacacagcatttgccttcgaa 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4547 AAAAGATGCTTTTCTGCTGAGACTCCACCCACGATAGTCTCTTCGCTTTGAG 4488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 79 tggagagatccagttacggaaagccgggacgtcacctggaacccgactgcccagaag 138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


D 4487 TGGAGAGATCCAGAGATGGGAATCTCAGACAAATTAACTGGACCAACTCCCGAGGAT 4428
Q 139 ttaagaactccccacatctttgacgaagccctaacagaggaacttggccaacttcaag 198
D 4427 TTCAAAACAGTCCACCTGTTGATGAAGCCCTGCACAGGAGACTGCAACACTTCGG 4368
Q 199 atcaaacacctcaagttagacctctccacagtagatgtagtgaacttcttcggcggaac 258
D 4367 ATCCAGACCCAGACCTGATTTGCTCAGTATGTAGTACTACTGCTGGCCGCACT 4308
Q 259 accaaacagactgcttagaagtagaaggaactacttgctgaattgtcgaactaagc 318
D 4307 TCTAGCTTACGTCTCAACAGGTACGGGGCCCTTTACAAACCTTAGGGACCTCGGA 4248
Q 319 taagagacctctgtgaagaagccagatttgcagagaagagtagtaacatacttgggttac 378
D 4247 TATGGGCTTCGGCCAGAAAGGCCAAATTGGCCAGAAAGGTCAAGTATCTGGGGTAT 4188
Q 379 agttgcgggacggcagcagatgctgcaggaagcaggaagaacatgtatgccagata 438
D 4187 CTTCAAAAGAGGTGACAGATGGCTGACGTAGAGGCCAGAAAGAGACTGTGATGGGGCAG 4128
Q 439 ccggcccccaacagcaagcaagtagaagatttltggggaacgtgatttgcaga 498
D 4127 CCTACTCCGAGACCCCTCCAGACTAAGGAGTCTTCTAGGAGAGGGAGGCTTCTGTCCG 4068
Q 499 ctgtgatcccggttltgcaccttagcagccccactaccgcgttaacaaagaataa 558
D 4067 CTTGTGATCCCGGGTTTGAGAAATGGCAGCCCTTGTACCTTCACCAAAAGGGG 4008
Q 559 gggggaattctctggtctccagacacagaagcatttgatctatacaaaagccctg 618
D 4007 ACTCTGTTTGAATGGGGCCAGACGACGAGAAAGGCTTACCAAGAGATCAAGCGCTCTC 3948
Q 619 ctgagcgacctgtctgtgacctccgcgaagtaactaaaccttaccctttagtggat 678
D 3947 TTAAGTGGCCCTGCTCCCTGGATGGCAGACTTGACTAAGCCCTTGCAGATTTTGTGTGAC 3888
Q 679 gagctlaaggaagtagcccgaggaattttaaaccacacttaggaacatgagagagact 738
D 3887 GAGAAAGCAGGCTACGCAAAAGGTGCTTACGCAAAAGATGGGGCTTGGCGGGCGC 3828
Q 739 gttgctactcttcaagaagcttbatctgttagcagttgttggccgtatgtctaaag 798
D 3827 GTGGGCTTACTGTCCAAAAGTTAGACCAAGTGGCAGCTGGGTGGCCCTTGGCTACGG 3768
Q 799 gctatcgacgtgtgacactagttgtaaggaagcgtgaacaaatgtaacttltggagcaat 858
D 3767 ATGTAGCAGCCATGCGCTTCTGACCAAAAGCGCTGGCAGCTCACCATGGCAGCCCA 3708
Q 859 ataactgtaataagcccccaatgcatltgagaaacatcgttggcagcccccaagcagatg 918
D 3707 CTAGTATTTCTGGCCCCCATGACAGTACGACATGTTAAAGCAACCCCTGATGGCTGG 3648
Q 919 atfacaagcccccgacatgacacatacaaaagcctgtcttct---cacagaaggttcaact 975
D 3647 CTCTCCAAAGCCCGAATGACCACTACAGGCTGTGCTTTCGACACGAGCGAGTCCAG 3588
Q 976 ttgcctcaccaagcgcgtctcaacctgtccactcttgcctggaagaactgtagaaca 1035
D 3587 TTGCGACCAATATGTCCTTAAACCAAGCTACGCTGCTCCCTCTA---CTGAGGAGGGG 3531
Q 1036 gtgacatcatgattgcatcaactatgattgtgaggaagcgtgggtctgcgaagaacttaca 1095
D 3530 CTGCAACATGACTGCTTGTACATCTTGGCTGAAGCCCAAGGAACTAGAACCAATCTTACG 3471
Q 1096 gacatacgccttgaatggaagatgtcactaactgttcaactgacgaagcagctatgtgtg 1155
D 3470 GACACGCTCTCCACAGCGCTGACACACCTGTGTACACAGATGGGAGCGCTTCTGTGAA 3411
Q 1156 gaaggtatgaagatgctgtggcggaatgtgtgacgaagcccgacgaatgttggcagc 1215
D 3410 GAGGGGACGCGCAAGGCGGAGCAAGTACAGTACCAAGCAAGCAAGTACTTCTGGGCCAAA 3351

Q 1216 agcttgcgggaagaaacttcaagcgcaaaagcttgaatcatalgaccttcaagcaagcttgg 1275
D 3350 GCACCTGCCACCGCGGACATCGGCCCAAGCGTGAAGTTGATAGCGCTCACCAACCTCTTA 3291
Q 1276 cggctgtgcgaaggaatacacaacttataaggaaggaagttagtgccttgcagct 1335
D 3290 AAATATGCGACAGATGAAGCTGAATGTTTACACGATAGCGCTTATGCTTTTCCACT 3231
Q 1336 gcaacgtacaagggccatctatacaaaaggggttcttaccctcaagcaggaagga 1395
D 3230 GCCCATTTCCGAGAAATATATAGAAAGCGCGGCTTCTCATCATCAAGAAAGAAAGAA 3171
Q 1396 ataagacaagaaggaatcttaagcctatitagaagccttaccatttgcacaaagctta 1455
D 3170 ATCAAAATATAGAGACGAGATCTTGCCCTTACTGAAGCTCTCTCTGCGCCAAAGACTT 3111
Q 1456 gctatitacaactgtctgtgacatcgaagaagccaagaatctctatcttagaaggaacag 1515
D 3110 AGCATTAATTTGCTCCCGGAGACATCAGAAAGGAAACCGCGGAGGCAAGGGGCAACAG 3051
Q 1516 atgctgtacccgggttgcgaagcagcgccagcgtgttaaccttctgctataatagaa 1575
D 3050 ATGGCCGACCAAGCGGCCCGAGAAAGTACCAAGAACTCCAGAGACTTCCACACTT 2991
Q 1576 acgcacaaagccccaagaagcc-agaagcagtagacaacctagaagacttgcgaagataa 1634
D 2990 CTGATAGAAATTTGACGCCCTTATACATCAAGCAATTTTCTACTAGCGGTGACTGACATA 2931
Q 1635 aaagatagacagacttctctgagactc-----cgaagggccgcgtctatacctatat 1686
D 2930 AAGATCTGACTTAACCTTAGGGGCCACTTATGACCATCAAGAAAGTGTGGTTATTCAG 2871
Q 1687 ggaaggaatactcgtccccaagaagggltagaatalgtlccaagataactcgtctata 1746
D 2870 GCAAGCGCTGTAATGCCATGATCAATTCACCTTGAACATATAGATTTTCTTCAATTAAG 2811
Q 1747 acccactaag-----aactaaacctgagaagttgtgtcgaatacctccatcat 1800
D 2810 ACCCAGCTCAGTTTCTCAAAAACAAAGGCTTCTTACAAAGAACTCTGTCTTATTTAC 2751
Q 1801 gttctgaggtacacagaggtgtgcgtgactcgtgtgtcaacaatltgttgccttgcagctg 1860
D 2750 ATGCTGAACCGGGATGCAAGCTCAAAAGACTCACTAGACTGATTCGCAAGCGTGGCACAG 2691
Q 1861 gttatgtctaatccttccaagaatactcccaagaagaagactaaggggaagccacccagc 1920
D 2690 GTCAATGCCAGCAAGCTGCGCTCAAAACAGGAGACTTGAAGTTGAGAGGCGCACCGCGC 2631
Q 1921 gctcactgggaagtgacttcaacttgaagtaagccggcttaatacgaagaacaaatcata 1980
D 2630 ACCCAGTGGGAAATGATTTCACTGAGGTAAACCTGGCTGTATGGGTATTAATATCTT 2571
Q 1981 ttggttltgtgaacaccttccaagatgggttagagcttactcctaagaagaagact 2040
D 2570 TTAGTTTTCATAGACACTTTCTCTGAGGTAAGACTTTCACCAACCAAGAAAGAACT 2511
Q 2041 tcaaccgtgtgtgtaagaataacttgaaggaatlttccaagaatttggaaatccaaag 2100
D 2510 GCCAAAGTTGTAAACCAAGACTACAGAGAAATCTTCCAGATTTGGCATGCGACAG 2451
Q 2101 gtaatgggtcagacaatgtgtccagcttctgttgcgaagtaagtcaaggaacttggcaag 2160
D 2450 GTATTGGGAACCAATAGGCGCTGCTGCTCCCAAGGTAAAGTCAAGAGTACGATGACGAT 2391
Q 2161 atattgggatttgaatggaactgtatgtgcatatacagaagcccaagctcaaggaagta 2220
D 2390 TTTATTTGGGGTGTATTTGGAATACATGTTGTCTACAGCCCAAGTTCAGGTGAGGTA 2331
Q 2221 gagaagatgataagacatlttaagaagcccttacaatlttaaccaaggaacttgcact 2280
D 2330 GAAAGATGATGATGAGCAATCAAGGAGACTTTTAACATAATTCACCGCTTGCAACTGCGCT 2271

OY	1099	ataccgcgtgacttggaagagatgcttaacatcttgcttcacgaaggaagcagctatgltggtga	1158
Db	5077	cagccgcctcccaagcgcgcgacacacccatggtglaacaagatggaagcagctctcttacaagag	5136
OY	1159	ggtlaagagatgctcttggtggtgagcagttgtgtgagaggaaccgcagcatctggccagcgc	1218
Db	5137	ggacacagctaaagtggtgagctgtgtgtgacacacgagaccgagtgatcttggcttaagcc	5156
OY	1219	ctgcgcgaaggaactcaagcgcgaagaagctgaatcctcagccctcaagcgaagcttgcgg	1278
Db	5197	ctgcgcacgcgcgacatccgtctcagcgggtcgtgaactgatagcatccacccaagccctaaag	5256
OY	1279	ctggccgaaggggaataatccataaacatttatcaggaagaagatgactgccttggcagctcga	1338
Db	5257	atggcagaagaagtaagaagcttaaaattgtttatctagatagaacggttatctgtcttcttcgc	5316
OY	1339	cacgtacagcgggcacatctataacaaagaaggggttcttacttaacgaagggaggggaata	1398
Db	5317	catatccatggagaataatactcaagaagcgltgggttgcctacatcagaagggccaagaagatc	5376
OY	1399	aagacaagaaggaatcttaagcctctatagaagccttcaattgtgcacaagaagctagct	1458
Db	5377	aaaataagaagcagatcttgcgcctataaagaagccctcttctgcaccaagaagactttagc	5436
OY	1459	attataacatgctctgtgacactcgcgaagcccaagaatctatctctagaagggacaagatg	1518
Db	5437	ataatccattgtctccagagacatccaagaaggaacacagctgcgcgaagcttagagagcaacg	5496
OY	1519	gtctaccgggttgctcaagcagcgcagcccaagcgtgtta-acctctcgtcctataatagaac	1577
Db	5497	gtctacaagcgcgcgcgaaagggagagccatccacagagaatccacgaacactccctctc	5556
OY	1578	gcccaaaagccccaagaacccaagacgaacatcacacccatagaagaatctgcagaagatataaa	1637
Db	5557	atagaaaattcatccatccctacacctcagccatccatcttcatcaacagtgactgatataag	5616
OY	1638	gataagacc-----agttcctgtgactccggagggagcactgctataactcatatggg	1699
Db	5617	gacctaaccaagttgtygggcattttatgtataaacaagaagtatgttgctctaccagaaga	5676
OY	1690	aaggaatccgtcccccacaagaaggttagatatgtctccacagatacatcgtcttaacc	1749
Db	5677	aaacctgtgactgtgacccagtttactcttggatatttagaactctctcaacagctgact	5736
OY	1750	caacct-aggaactaacaacctgcagcagtttgcagaaatctcccttcatgttcttag	1808
Db	5737	caacctcagcttccaaataatgaaggtctctccatagaagagagccacagctccctactaatg	5796
OY	1809	gtctccagagatgctgctgactcgtgtgtgtacaacatgtgtgtccctgcagctggttaatgc	1868
Db	5797	ctgaacccgggatactgaataactcaaaaatactactgtgaacccctgaagaagcttgtgcaaatgc	5856
OY	1869	taacctctccagaatatactccaggaagaagcactaaag-----ggaaagccaccagcgct	1923
Db	5857	aagcgcgcgaagctcgtccgtttaaacaagggaaactaaggtgtccgcgggcatacgcgcgcgtc	5916
OY	1924	caactgggaagtgtgactcactcgt	1983
Db	5917	catgtggagatgatcttcaacgcgataaagcccgatgtgtatgtgtatataatctctcta	5976
OY	1984	gttttttgtaaacaccttttcaagatgtgtgtatagaggtctatcctactaagaagaagacttca	2043
Db	5977	gttttttataatctcttcttctgtgtgtgtatagaagccttcccaacaaagaagaacgcgc	6036
OY	2044	accgtgtgtgtctagaanaaaactcgtgggggaatttttccaagaatttggaatacctaaagta	2103
Db	6037	aaggtcttaaccgaagaagctactcagaaggagatcttccccaaggttttcggatccttcaagta	6096
OY	2104	ataaggtctcaacaatgtgtccagcttctcgttccgcgaagtgaagtccagggatctggccaagta	2163
Db	6097	ttgggaacttgacaatggtgcctgctctcgtctcccaagtgatgtgcagaagatgtgcgcgactgt	6156
OY	2164	ttgggagttatgtggaactgtcatgtgtgtcatcaagacccccaagctcagaagcaggtlagag	2223

[illegible]

```

RESULT 14
US-09-011-745-4
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Lolc
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: Construct
US-09-011-745-4

```

[illegible]

Dh 6217 agaataagatagaacacatcaagaagacttaactaaattaacgcttgcaactggtctaga 6276
Qy 2284 gatgagtggtctcccgcccttctgcttttaaggtaggaacaccccttgacaagttt 2343
Db 6277 gactgggtgctccttaacccctttagcccttgtaacagacccgcaacgcccggccccc 6336
Qy 2344 gggctggcccttaagttgctctacgggggaccccccccgcttggcaagaattgcttt 2403
Db 6337 ggcctcaacccatataatgatactatataatgggcaaccccccccgcttgaattccctgac 6396
Qy 2404 gcacatagtgctgtagtgcgtcttcccaagccttcttctctagctcgaagcgctcag 2463
Db 6397 cctgacatgacaagaagtacttaacagccctctcacaagctcaacttaacaggtctctac 6456
Qy 2464 tgggtgagcagcagcggttggaagcagctccggagagccctacccaagaagaagcttgcaa 2523
Db 6457 ttagtcagcagcagcagcttggaagcagctccggagagccctacccaagaagaagcttgcaa 6516
Qy 2524 -----gtccacatgccttcccaagttggagatcagtcatalgttagacgcacgtgca 2577
Db 6517 ccggtgtactcctcaaccccttaacgagtcgagcaacagtggtgctcgcgcaacacagact 6576
Qy 2578 ggaacactcgagactcggttggaagggaaccttactctgactttgaccacacacagcgt 2637
Db 6577 aagaactagacactcgcttggaagggaaccttactctgactttgaccacacacagcgt 6636
Qy 2638 gtgaagtcgaagaatcccaactgatacagatccacagcttaagccggc 2690
Db 6637 ctcaaatgagcagcagcagcttgatgatacagcgcgcaacgtagaagctgc 6689

RESULT 15

US-09-011-745-2
: Sequence 2, Application US/09011745
: Patent No. 6165715
: GENERAL INFORMATION:
: APPLICANT: Collins, Mary KL
: APPLICANT: Weiss, Robin A
: APPLICANT: Takeuchi, Yasuhiro
: APPLICANT: Cosset, Francois-Joic
: TITLE OF INVENTION: Expression systems
: FILE REFERENCE: 09/011,745
: CURRENT APPLICATION NUMBER: US/09/011,745
: EARLIER FILING DATE: 1998-06-22
: EARLIER APPLICATION NUMBER: PCT/GB96/02061
: EARLIER FILING DATE: 1996-08-23
: EARLIER APPLICATION NUMBER: GB9517263.1
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 2
: LENGTH: 7616
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Portion of
: OTHER INFORMATION: construct
US-09-011-745-2

Query Match 27.6%; Score 916.2; DB 4; Length 7616;
Best Local Similarity 60.6%; Pred. No. 2.3e-268;
Matches 1633; Conservative 0; Mismatches 1033; Indels 27; Gaps 7;
Qy 22 gatgccttctctgctgagattaacacccactagccaacacacttttgccttcogaatg 81
Db 3161 gatgccttctctgctgagattaacacccactagccaacacacttttgccttcogaatg 3220
Qy 82 agaatacaggtacgggaagaacacggcagctcactcgccgacacggcagcagggtt 141
Db 3221 agaatccagatgaggaatcctcaggaactgacccctgagacacccccaagggttcc 3280
Qy 142 aagaactccagcactcttgacgaagccctacacagggacgtgccaacttcagatc 201

Dh 3281 aaaaacagcccaaccccttgatgagcactgcacagacactagcagacttccgagtc 3340
Qy 202 caaacacctcaagtgaaacctctcctcagtaacgtgtagacctctctgctgggagcaacc 261
Db 3341 cagacacacagactgactctgctacagtaacgtgtagacctctctgctgggagcaacctct 3400
Qy 262 aacacagactgcttaaggaagtgacgaagacactgcctggaattgtcgaactaagctac 321
Db 3401 gagctagactgcacaaaggtactcgggcccgttacaacaccttgggaactcgtgact 3460
Qy 322 agagcctctgctcaagaagcccaagatttgcaggaagaggttaacacttgggttagact 381
Db 3461 cgggctctggccaagaagcccaatttgcagaaacaggtcgaagtatcttgggtatctt 3520
Qy 382 tgcgggagcggcagcagatgctgacgagagcagcaggaagaacactgtatccagttacg 441
Db 3521 ctcaaaagaggtctagagatgctgactgagcccaaaagagactgtatgggagacct 3580
Qy 442 gcccaacacacagccaacagtgagagagtttgggagacagctgatttgcagactg 501
Db 3581 actcgaagaacccctcgaacactaagggatctcctagggagcagcagcttctgctcctc 3640
Qy 502 tggatccggggttgcagacttgcagagcccaactcactacccgttaacccaagaagaagg 561
Db 3641 tggatccctgggttgcagaaatgagccccccttgtaacctctcacaacaaacgggagact 3700
Qy 562 ggaattcctctggctctgacgacacgaagagatttgaatgataaaaaggccctgctg 621
Db 3701 ctgttaatttggggcccaacccaagaagggccttcaagaataaagaacactctctca 3760
Qy 622 agcgacactgctgctgcccctcctcagcgaactaaaccccttaaccttatgtgatgag 681
Db 3761 actgcccagacccctggggttgcagacttgcagacttgcagacttcttgcagcag 3820
Qy 682 cgtaaaggagtagcccgaggaatttaacccaacccctaggaacataggaagagactgt 741
Db 3821 aagcagggtagcgaagaagtgctcctaaacgaacaaacttggacacttgcgtgcggtg 3880
Qy 742 gcttactctcaaaagagcttgactcctgtagcagagtggttgcgtagcttgaagct 801
Db 3881 gcttactctcaaaagagcttgactcctgtagcagagtggttgcgtagcttgaagct 3940
Qy 802 atgcagcgtgtagccatctgtagcagagcgtgcaacaaatltacttgggaacagata 861
Db 3941 gtagcagcattgctcgtactgacaaagagtgtagcgaacaaatltacttgggaacagata 4000
Qy 862 actgtaataagcccccaatgaggaacatcgttgcgagcccccagacagatgagatg 921
Db 4001 gtagcttgcgcccccatgtagtagagcactagtagcaaacacccccagacgctgact 4060
Qy 922 accaagcccgatagcccaactataaagcgtcttct--cacagagaggttactctc 978
Db 4061 tccaagcccgatagcccaactataaagcgtcttcttgggaacaggaacgggttccagttc 4120
Qy 979 gttccacacgcccctcacaacccctgacactcttctgctgaagagactatgaacacagt 1038
Db 4121 ggaacggtgtagccttgaaccccgtagcgtctctcc--actgcttgaaggaaggtctg 4177
Qy 1039 actcatgattgcataactatgattgtaggaagcagctgggtccgaagaacactacagac 1098
Db 4178 caacacaactgcttgalatcctgtagcgaagcccaagcaggaaccccgacacttaacgagc 4237
Qy 1099 ataccgctgacttgaagaagtgtcaacttggttcactgacgaagcagcagctatgtgtgaa 1158
Db 4238 cagcgcctcccaagcgcgaacacacttgttacaaggatggaagcagctcttcaacaagg 4297
Qy 1159 ggttaagagatgctgtggtggcggcagtggtggaacgggaacccgacagatctggtccagcagc 1218
Db 4298 ggaacagcgttaagcgggagctggtgtagcaccaggaacccgaggtatctcgtgttaagcc 4357
Qy 1219 ctgcggaaggaactcaagcgaagaagcgtgactatgagccctcaacgcaagcttgcgg 1278

Db 4358 ctgcacgacgagacatccgctcagcggtcgaatgatalagacatccacgagccctaaag 4417
 QY 1279 ctgcccgaaggagaaatccatlaaacaattlaacggaacaggtatgctcttgcagctga 1338
 Db 4418 atgcaagaagatgaagaagctaaatgtttatctatgataagcgcttatgcttctgctgc 4477
 QY 1339 caacgtacacggggccatctataacaagaagggtgttgctacctcagcagagaggaataa 1398
 Db 4478 catatccatggaataatatacagaagcggtgtgtgtctcacatcagaagcagaagatlc 4537
 QY 1399 aagaacaaggagaaatcttaagcctattagaagccttaacatttgcacaaaggtatgt 1458
 Db 4538 aaaaaaagaacgagactcttgccctacaaagccctcttctcccaagacttgc 4597
 QY 1459 attatacactgtcctcgtacatcagaagccaagatcctacatactcagaagagacagatg 1518
 Db 4598 ataatccatgtccaggaatcaaaaaggacacagcgccgaggtcagagccacggatg 4657
 QY 1519 gctgacgggtgtgccaagcagcagccaggtgtta-accttctgtctataatagaac 1577
 Db 4658 gctgacaaagcgccgaaagcagccatcagaagactccagacacccctcctcctc 4717
 QY 1578 gcccaaaagcccaagaacccagacgacagatcacccctagaagacttgcagagataaaa 1637
 Db 4718 atagaaaatltcatccctacacccctcagacatttcaatcacagtgactgataaaag 4777
 QY 1638 gatagaac-----agttctgagactccggagggagccgtatacctacatattgg 1689
 Db 4778 gacctaacaagaattggtggccattcatgataaacaagaatattggtcttaacaagga 4837
 QY 1690 aaggaatctctcccaacaagaaggttagaatagttccacacagataatcgtctaac 1749
 Db 4838 aaacctgtatgtcgtacacagtttaacttgaatttagacttcttcaatcagctgact 4897
 QY 1750 caact-aggaataaacctctcagcagttgttcagaacatcccttcatgttctgag 1808
 Db 4898 caactcagctctcacaagaatgaaagctcctcctagaagaagccacagtcctcactacatg 4957
 QY 1809 gctacacagagatgctgactcgtgtgtaacaacattgtgtccctgcagctgttaatgc 1868
 Db 4958 ctgaacccgggactcgaacatcaaaaatactgaacttgaaagcttgtgcacaagtc 5017
 QY 1869 taatcctccagaataactccacaggaagagactaag-----ggaagccaacccagcgct 1923
 Db 5018 aacgcacagcaagctcgtccgttaacaaggaactaggtcgcggcagcagccgcact 5077
 QY 1924 cactgggaagtgaacttaactgaagcgggttaataatcagaaacaatatctatgtg 1983
 Db 5078 catgtggagatcgaatltcacccgagataaagccggatgtgtatgctataaataatcttcta 5137
 QY 1984 gttttgtagacacttttcaagatggttagaggttactcactaagaagaagacttca 2043
 Db 5138 gttttatagataccttttctgctgctgataagaccttcccaacaaagaagaacggc 5197
 QY 2044 accgtgtgtgctaaagaataactcaggaataatttccaagatttgaataactaagta 2103
 Db 5198 aaggtcgtlaaccaagaactactagaagagatcttcccaagttcgcgactcctcagta 5257
 QY 2104 atagggtcagaaatggtccagcttctgtgcccaggtaaagtcagggagctgccaagata 2163
 Db 5258 ttgggaacgtgaactcgtccgtcgtccaggtcaggtcagacagtgccgacttg 5317
 QY 2164 ttgggagatgtgaactcgtatgtcatatcagaccccaagctcagagcaggtagag 2223
 Db 5318 ttgggagatgtgaactatcatatgtgcatatcagaccccaagctcagagcaggtagaa 5377
 QY 2224 aggaatgaatagaacatlaaagagaccccttaacaaatgacacagagactgtgatalat 2283
 Db 5378 agaatgaatagaacatcagaagagacttaactaataatgaacgttgcacactgtcctaga 5437
 QY 2284 gatlgagtgactcctcgtcctgtgtgttttttaggtgaggaacacccctgagacattt 2343
 Db 5438 gactgtgtgtcctactccttgaacgttgaacgagccgcaacacgcccggcccat 5497

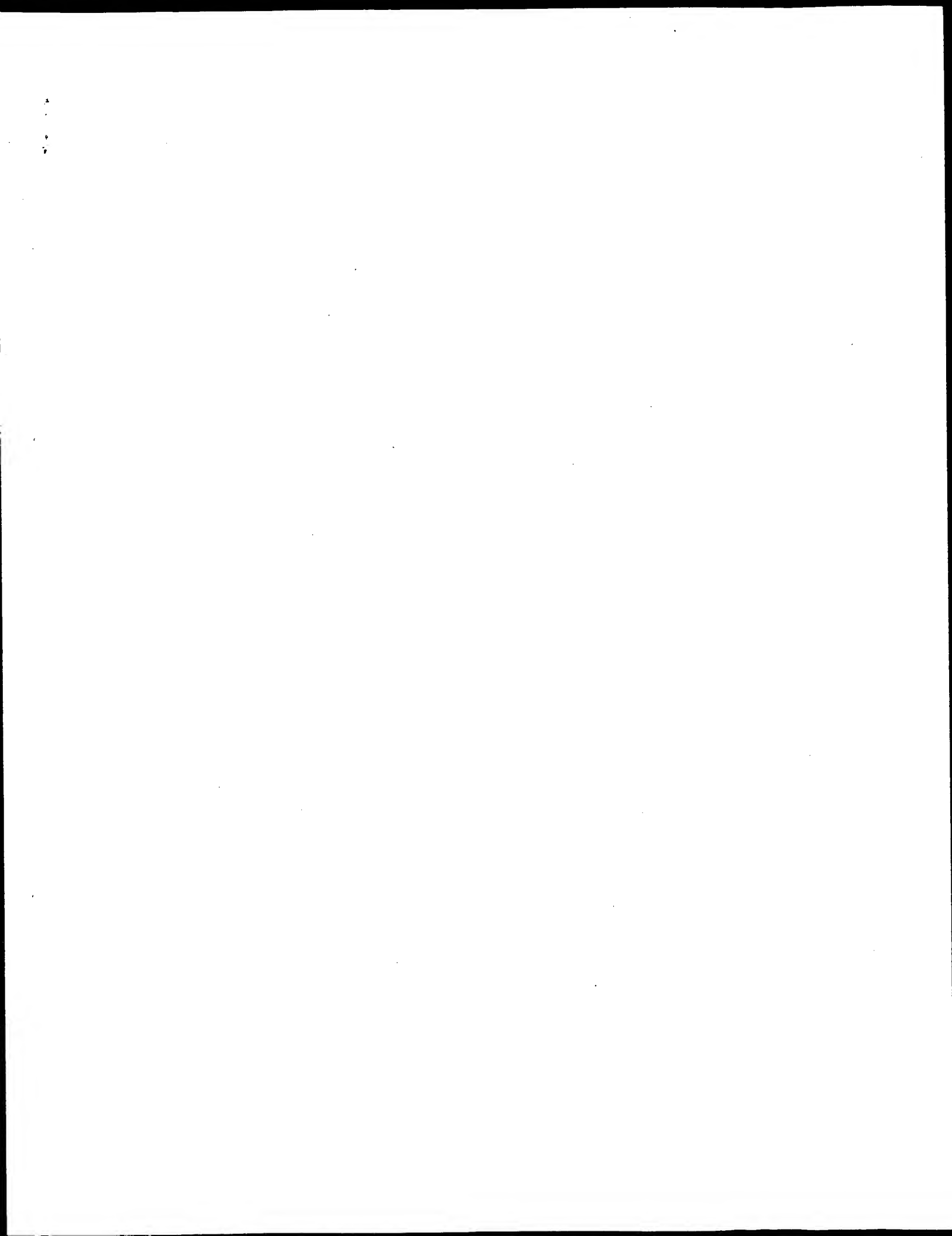
QY 2344 ggcctgaccccatatgaattgtctacggggggaaccccccggtgagagaattgcctt 2403
 Db 5498 ggcctacaccatattgaattcttatatggggaccccccggttgaattcctccgac 5557
 QY 2404 gcaatagtgctatgtgtgtgtcttccagccttgttctcctaggtcagaagcgctcag 2463
 Db 5558 cctgacatgaagaagattactaacaagccctctcctcaagctcactacaggtcctctac 5617
 QY 2464 tgggtgagcagcagcggtggaagcagctccggagccctcactcagaaggaagacttgcaa 2523
 Db 5618 ttagtccagcagcaggtcgtgagacctctggtgagccttacaagaacaactgagacga 5677
 QY 2524 -----gtccacatcgtcctccaaqtttgagatcagctcatgttlaagcgcacgtgca 2577
 Db 5678 ccggtgtgtactcaacccttaaccgagctgcgcagacaggtgtgtcgcgcgacacagact 5737
 QY 2578 ggaacctcagagactcgtgtggaaggaacctatctcgtactttgaccacacaagcgct 2637
 Db 5738 aagaacctgaacctcgtgtgaaagaccttaacacagctcctgtgacacccccacggc 5797
 QY 2638 gtgaagtcgaagaatcccaacctgtgatacagatccacagttaaagccgc 2690
 Db 5798 ctcaagtaagcagcagctcaggttggatatacagcgcccaagctgaaggtgc 5850

Search completed: February 24, 2002, 02:39:30
 Job time: 20121 sec

Mon Feb 25 07:43:58 2002

us-09-171-553b-1.feb22std.rml

Page 25



4373 ATTAGAGCCGTACATTTCACAAAAAGCTAGCTTTATACACTGTCTGACATTCAGAA 4432
4436 agccaagaatcatalatcagaaggaaccagatggctgaccgggttgccaagcagcgc 4495
4433 AGCTAAAGATCTCATATCCAGAGAAACAGATGGCTGACGGGTGGCAAGCGAGC 4492
4496 ccaagctgttaacctcttgccctataatagaacgcgaagcccaagcccaagagaca 4555
4493 CCAGGCTGTTAACTTGTGCTTATATAGAAATGCCAAAGCCCGAGAACCCACACACA 4552
4556 gtacaccctagaagatgcaagagataaaaga tagaaccagtctctgagaccgga 4615
4553 GTACACCCCTAGAAAGATGGCAAGATAAAAAGATAGACAGTTCTCTGAGACTCCGGA 4612
4616 ggggacctgctatacctaatatggaagaaatcctgcccacaaagaaggttagata 4675
4613 AGGGACCTGCTATACCTCAGATGGAGAAATCTGTGCCCAAAAGAGGGTTAGATA 4672
4676 tgtccaagatatacctgcttaaccacactagaactaaacaactgacagcttggtcag 4735
4673 TGTCCACAGATACATCGCTTAACCCACTAGGAATTAACACCTGCAGCAGTTGCTCAG 4732
4736 aacatccctatcatgtctcagagctaccaggaatggtgctgactggtgtcaaatgt 4795
4733 AACATCCCTTATCATGTTCTGTAGGCTACACAGAGTGGCTGACTGCTGCTCAAACTTG 4792
4796 tctgacctgacagctggttaatgtaactcttcacgaactctccagaaagagataag 4855
4793 TGTGCTGCTGCACTGCTTAATGCTTAATCTTCAGAAATGCTCCAGGGAAGAGCTAAG 4852
4856 ggggaagccacccagcgtcactcaggaagtgactcactgaggttaagccgcttaata 4915
4853 GGGAGGCAACCCAGCGCTCAGCTGAGAACTGAGACTGAGTAAAGCCGGCTAAATA 4912
4916 cggaaataatactatgttgttgttgaacacttcaagaatggttagaagctatcc 4975
4913 CGGAACAAATACCTATGTTGTTTGTAGACACTTTTGAAGATGGGTAGAGCTTATCC 4972
4976 tactaagaagaagactcaaacgttggttgcttaagaataactcgtgagaaattttccag 5035
4973 TACTAAGAAAGAGCTTCAACCGTGGTGCTAAATAAAATCTGGAAGAAATTTTCCAG 5032
5036 atttggaatacctaaggttaaggtcagacaaatggttccagcttccgttggccaagtag 5095
5033 ATTGGAAATACCTAAGGTATAGGGTACAGCAATGGTCCAGCTTTGTTGCTCCAGAGTAG 5092
5096 tcaaggaactgccaagatatctggggaatgtaatgaaactgcatgtgtcatatagaccca 5155
5093 TCAAGGACTGGCCAAAGTATTTGGGATTTGGAACACTGANTGTGATACAGACCCCA 5152
5156 aagctcagaagtagagaagatgaatagaacatlaaagagaaccttaacaatgtac 5215
5153 AAGCTCAGAGACAGGTAGAGGATGAATAGAACCATTAAGAGAACCTTACTAATGAGAC 5212
5216 cagcgaactggaataatgaatgaatgagctcctcgccttctgtctttagaggttag 5275
5213 CGCGAAGACTGGCTTATATATTTGGATAGCTTCTCTGCTTGTGCTTTTAAAGGTTAG 5272
5276 gaacaccttggaagcttgggctggaacctta tgaattgctctactagggggaacctcccc 5335
5273 GAACACCCCTGAGACAGTTTGGGCTGACCCCTATGAATTACTCTACGGGAGACCCCCC 5332
5336 gttggcagaagaattgctcttgacatagtgctgagtgtgctgttccccagcttgtctc 5395
5333 ATTGGTAGAAATTTCTTCTTACATATGCTGACGTGCTGCTTCTTCCAGCTTTGTTTC 5392
5396 taggtcaagcgtctcagatggtggtgagcagcagcgttgaagcagcctccggagagccta 5455
5393 TAGGCTCAAGGCTTGTGAGGTGAGACAAAGACGCTGAGGCAACTCCGGAGGCTTA 5452
5456 ct 5512
5453 CTCAGAGAGAGAGACTTGGAGATCCACATCGTTTCCAGTGGAGATTCAGTACGT 5512

5513 tagagccaacgttagaagaaactcgaagactcgttggaaggaactatctctactt 5572
5513 TAGAGCCACCCCTGAGAGAAACCTCAGACACTGGTGAAGGGCCCTTATCTCGACTTTT 5572
5573 gaccacacaagcgtgtgaaagtcgaagaaatcccaactgataatgcatccagct 5632
5573 GACCACACCAACGGCTGTGAAGATCGAAGGATCTCCACCTGATCATCATCCACGT 5632
5633 taagccgggagcactcccgactcgggttggaagccgaagaactgatacccttaa 5692
5633 TAAACCGGCGCACCTCCGATTCGGGTGGAAGACCAAAAGCTGAATAATCCCTTAA 5692
5693 gcttcgctcaactcgtgtgtctctacttaacaataactcccaagcagtagtaaac 5752
5693 GCTGCGCTCCATCCGCGGTGCTTCTTACTGTGCAATTAACCTCAGACATAAGGTATGC 5752
5753 gacctatagaagctcgaacccccctagaacttataccctactctgctgataatgac 5812
5753 GATAGAGACAGCTGAACTCCATTAACCTTATCTCTCACTGCTTAATTAATCTGACT 5812
5813 ctatagcgtgtcaactgtaaatagcactcgaagtgtgtctcctagaagcactgtgac 5872
5813 CCGGACAGATTTATATATCAACACACCAAGGAGGCTCCTTAGAAGACTGTGGC 5872
5873 ctgaactgacttctgctcctcgaattgataaccgcgtgataaagcactctccaac 5932
5873 CTGATCTATACGTTTGGCTCAGATCATGTTATCTTATGTA--CTCACCCCAAGATA 5929
5933 tagtccgtatgtagtggtctctatctgtccagcagcagc---agaagaagaatactgtg 5989
5930 TCTTCATGCTCAGAGATTTTATGTTTGGCCAGGACACCAAAATATGAAACATTGGC 5989
5990 ggggtctcgggaatcctctcgaagatgagagctgcgtcaactcacaagatgagact 6049
5990 GAATCCACAGATTTCTTTTGAACAAATGAACTGTGTAACCTTATATGATGATTT 6049
6050 ggaatggccagatctctccaggaacgggtgaatctccttctca----- 6098
6050 GGAATGGCCAACTCTCAGCAGAGATAGGATAGGTTTCTTATTCACACCTATACCA 6109
6099 -ttccgcccggcgaagatacaaatgataatgataatagaatagagctgtcccat 6157
6110 GCTCTGACAAATTTATTTACTGACCTGATTAAGAACTGAGAGCCCAAGTGTCTCTT 6169
6158 cagaactagaattatcaagaatgaattcact--gaagaagaacaggaataatcaaa 6215
6170 CAGACTTATGATCTTAAATATTAATTTCACTGAGAAAGGAAACAAATATCTTAA 6229
6216 agtgataaattgtagagctggaagatagttttatataatgagcgtg---ggaagcag 6271
6230 AATGGATTAATGATATGTCTTGGGAATGCTATATATGAGAGCTCGGTAACAACAG 6289
6272 gttcacttaacatctgctttagatagagaaggggacaaacccctgtggaagcgg 6331
6290 GCTTCATTTATACCTTATGCTTCAAAATTAACCAAGTGTG--GAGCTTCAATAGCTTAG 6346
6332 gaccgataaagtaactgctggaagcaggggccccggcctgagcacaactgtgc 6391
6347 GACCAATATGAGTGTGACGGGTCAAGAGACCCCAACCC----- 6385
6392 cgtgtcccaataactcgtcgtgagcactgacataaacaagcgccttagaagtagta 6451
6386 -----AAGGACCAAGGACCATCTCTTACATTAACCTT 6415
6452 ctgattgattcctacacaacagcctagaacatcccccaggtgttctgtttagaagcag 6511
6416 CTGATAGAGACCCACTGAGCTTACAGCAGCAGC-----TAAATAGGGG 6460
6512 agaaactctcagctatcaggaagccttcaagccatcaactcaccgagcctgagt 6571
6461 CAAACTTTTATGCTTATCCAGAGAGACTTTTAACTCTTAACCTCAGACTCCAGAG 6520

QY	1980	agagagagagatttaggaaatagtcagccctagacagtcaggaacctgagcaat	2039	QY	3060	caatcaccctggaccgactgccccaagggttcaagaactcccccagaccatttcagaa	3119
Db	1990	AGAGAGAGAGATTAGAAAAATTAGTCAGGCCCTTACAGCTCAGGAACTGGGCAT	2049	Db	3070	CAGGTTACCTGGAGACCCGACCTGCCCAAGGGTTCAAGACTCCCGACATCTTTGACGAA	3129
QY	2040	aggaccccaactcgcacaaggaccagtgtgcattgttaaaaaaagcaacttggcaga	2099	QY	3120	ggcctacagaaggacctgcccnaacttcaagatcccaacccctaagtgtacccctccag	3179
Db	2050	AGGACCCCACTCGACAAAGGACAGTGTGCTATTGTAAGAAAAAGACACTGGGCAAG	2109	Db	3130	GCCCTACACAGGAGCACTGGCCAACTTCAGATCCAAACACCTCAGGACTCTCTCCAG	3189
QY	2100	aactcccccaagaagaagaagaaggaccgaaggtccctagctctagaagaagataagat	2159	QY	3180	taagtgtatgacctgtcttcgtgcggagagcccaaaagagactgtctagaagtagaag	3239
Db	2110	AACCTCCCCAAGAAAGGAAACAAAGACGGAAGGTCTTAGCTTGAAGAAAGATAAAAAT	2169	Db	3190	TACGTGATGACCTGCTTCTGGGGAGCCACCAAAACAGAGACTGTAGAAAGTTAGCAAG	3249
QY	2160	taaggagagaggggttcggacccccctcccgagcccaaggttaactttaaagttgaagg	2219	QY	3240	gcaactatgcttggaattgtctgaacctagagctatagaagacctctagaaggccagatt	3299
Db	2170	TAGGGGACACGGGGTTCGGACCCCTCCCGACGCCAGGATTAACCTTGAAGTGGAGGG	2229	Db	3250	GCACTTACTGCTGCAATTTGCTGACCTAGCGTTACAGACCTCTCTTAAGAAAGGCCAGATT	3309
QY	2220	caaccagttgagttccgtgttgataccggagcggagacatcagtgctgtacaacatla	2279	QY	3300	tgcaaggagagagttacaatacttggggttacagtttcggggcgaggcagatgtgtgacg	3359
Db	2230	CACCAGTTGAGTTCTCTGTTGATACCGAGCGGAGCATTCAGTCTCTACAAACATTA	2289	Db	3310	TGCAGGAGAGAGTAACATTACTTGGGGTACAGTTTGGGGGGGGGAGCATGTGACG	3369
QY	2280	ggaacttaaaaaaagaaatccctgggtgtatgggtgcacagggcacaagtcagatcca	2339	QY	3360	gaaggcaaggaaagaactgtatgtccagataccggcccccaaccccaagcaagtgaga	3419
Db	2290	GGAAACTAAAGAAAAAATCCGGGTGATGGGTGCACAGGGCAACGGCAGTATCCA	2349	Db	3370	GAGGCACGGAAGAAAACTGTAGTCAGATVCCGGCCCAACACACGCAAAAGTGAGA	3429
QY	2340	tggaactaccggaagaccgttgacttggagtggaacgggttaacccactgttctgtgc	2399	QY	3420	gaatttttgggaagcgttgattttgaagactgtgataccggggttgcagacttagca	3479
Db	2350	TGGACTACCCGGAAGAACGTTGACTTGGAGTGGGACGGGTAAACCACTCTTCTGTGTC	2409	Db	3430	GAGTTTTTGGGACAGCTGATTTTTCAGACTGTGATCCGSGGTTTGCAGACTTAGCA	3489
QY	2400	atccctgagtgcccaagtaeccctctagtagagagacttaactgaacaaatgysagctoa	2459	QY	3480	ggcccaactcaaccgcgttaaccaaagaagaaggggagatctcctggctcctgagcaag	3539
Db	2410	ATCCCTGAGTGCCTCCCTTCTAGTAGAGACTTACGACCAAGATGGAGCTCAAA	2469	Db	3490	GCCCCACTCTACCCGCTTAACCAAGAA-----	3516
QY	2460	attcttttgacaagaagagaccgaaggtgtctgtgaaatacaaacatctactgtttg	2519	QY	3540	aaggcatltgatgtctalcaaaaaagccctgtgtgagcgcaactgtctgtgcccctctgac	3599
Db	2470	ATTCTTTTGAACAGAGAAAGACCAAGATGTCTGTGAATTAACAACCATACTGTGTG	2529	Db	3517	-----	3516
QY	2520	aacctcaatlagatgatgaatcgcataatctccccaagtaaaagcctgtacaagat	2579	QY	3600	gtactaaacccttlaaccttlatgtgtagagcgtlaaggagtagccgagagtitta	3659
Db	2530	AACCCCAATTAGATGATGATATCGACTATTTCTCCCAAGTAAAGCCGTGATCAAGAT	2589	Db	3517	-----	3516
QY	2580	ataagtcctgtgttggagcaatttccccaagccttgggacagaacccgagagatggtttg	2639	QY	3660	accocaaacctagagacacatgagagagacctgtgtcctactcgtlcaagaagcttgact	3719
Db	2590	ATAAGTCTCTGTGGACCAATTTTCCCAAGCTGGGCGAGAAACCGAGGATGGTTTG	2649	Db	3517	-----	3516
QY	2640	gcaaaagcaagttcccccacaggttatcaactgaagccaggtgtacccaagatcagtc	2699	QY	3720	gtagccagtggtggcccgatgtctgaaggtcatcgcagctgtgtgcatactgtgcaag	3779
Db	2650	GCAAAAGCAAGTCCCCACAGGTTATTCATCTGAAGGCCAGTGTACACAGTATCAGTC	2709	Db	3517	-----	3516
QY	2700	agacagtaecccttgagtagagagagcctgagaagaaatttgccgcagtgtlcaagaata	2759	QY	3780	gaagctgacaatltgactlttgagacagaalatalactgtaagcccccatgtatggag	3839
Db	2710	AGACAGTACCCCTTGAGTAGAGAGGCTCGAGAGGAATTTTGGCGCAATGTTCAAGATTA	2769	Db	3517	-----	3516
QY	2760	atccaacaggagcatcctcagttcctgtccaatcccccttggaaactccccctgtacaggtt	2819	QY	3840	aacatggttcggagcacccccagacagctgtagtgaaccaagcccgagtagccaactatcaa	3899
Db	2770	ATCCAACAGGGGCAATCTAGTTCCTGTCCTCAATCCCTTGGAAATATCTCCCTGTAACCGTT	2829	Db	3517	-----	3516
QY	2820	aggaagccttgagaccaatgatattcgaaccagtagcaggaacttgaagaggttcaataaag	2879	QY	3900	agcctgtcttcaacagagaggtgtactcttcgtcccaacagccgcttcaacctgtccact	3959
Db	2830	AGGAAGCCTGGGACCAATGATTTATCGACATACAGGACTTGAAGAGAGTCAATTAAGAAG	2889	Db	3517	-----	3516
QY	2880	gtgcagagcatatacccaacgggtcccgaaacctataaccttgaagcgccctccgct	2939	QY	3960	cttctgcttgaagagactgtatgaaccaagtgtactcatgaltgtcoactaatatgttag	4019
Db	2890	GTGCAGGACATACACCAACGCTCCGAAACCCCTTAACCTTGAAGCGGCGCTCCGCT	2949	Db	3517	-----	3516
QY	2940	gaacggaactgttacacagtagttgactttaaagatacctctctcctgcctgagattacac	2999	QY	4020	gagactggggttccgcaagaaactttagacataacgctgtcgtgaggaagtgtctaacttgg	4079
Db	2950	GAAAGGAACTGTACACAGTATTTGACTTAAAGATGCTTCTTCTGCTGAGATTACAC	3009	Db	3517	-----	3516
QY	3000	cccaactagccaacacttltgtccttcgaatgtgagagatccaagttaagtggaagaccggg	3059	QY	4080	ttcaactgaggaagcagctatgtgtgtgaaggtlaagagtagtctggggcggtcagltgtg	4139
Db	3010	CCCACTAGCCAAACCTTTTACTCTGAAATGGAGAGATCCAGGTACGGGAGAACCCGGG	3069	Db	3517	-----	3516
QY	4140	gacgggaccccgacgaltctgggcagcagcgtgcgggaaggaacttcagcgcaaaagct	4199	QY	4140	gacgggaccccgacgaltctgggcagcagcgtgcgggaaggaacttcagcgcaaaagct	4199

Db 3517 ----- 3516
Oy 4200 gaagctatgcccctcaagcaagcttgcgctgcgcaaggaatacctaataattat 4259
Db 3517 ----- 3516
Oy 4260 aagcaagcaggtatgcttgcagctgacagctacagggccatctataacaag 4319
Db 3517 ----- 3516
Oy 4320 gggcttcttaccagcaaggggaaataaagaacaagaagaattcttaagctta 4379
Db 3522 ggggtgcttaccctgagcaggggaaataaagaacaagaagaattcttaagctta 3581
Oy 4380 gaagccttaccattgcccagaagaagctgctatatacaactgctctgacatcaagaagc 4439
Db 3582 GAAGCCTTACATTTGCCAAAAAGGCTACCTATTTACACTGTCTGTGACATCAAGAAGCC 3641
Oy 4440 aaagctcatatctagaagggaacagatgctgacggggttgcgaagcagcagccag 4499
Db 3642 AAAGATCTCATATCTAGAGGGAACCAAGTGCACCGGGTGGCCAAAGCAAGCCAG 3701
Oy 4500 gctgttaacctctgctataatagaaacgcccagaagcccagaacccagcaat 4559
Db 3702 GCTGTAACTTCTGCTATATATAGAAAGGCCAAAGGCCCAAGCCAGACAGTAC 3761
Oy 4550 accctagaagactgcaagagataaagaatagaccagttcttgcagctccggaagg 4619
Db 3762 ACCCTAGAAGACTGCAAGAGATAAAAAGATAGACCAAGTCTGTGAGACTCCGAGGG 3821
Oy 4620 acctgctacatctatagggaagaatctctgcccagaagaagggttaagatctc 4679
Db 3822 ACCTGCTATATCTATAGGGAAGGAATCTCTGCCCAAGAAAGGTTAAGATATGTC 3881
Oy 4680 caacagatatacgtctcaaccaccatagaaactaaacactgcagcagttgctaga 4739
Db 3882 CAACGATACATCGCTAACCCACCTAGAGACTAACACCTGCAGCAGCTTGCTGAGAA 3941
Oy 4740 tcccttatacttcttgagagctccagagtgctgactgctgctgctgctgctgctg 4799
Db 3942 TCCCTTATCATGTTCTGAGGCTACCCAGAGAGTGCCTGACTGCTGCTGCTGCTGCTG 4001
Oy 4800 cccctgcaagctggttaagctcaatccctccagaatccctccagaagaagaagagga 4859
Db 4002 CCGTCCAGCTGCTTAATGCTAATCTCTCCAGAACTCCAGGAAGAGACTAAGGGA 4061
Oy 4860 agcaaccagcagctcactggaagtggaactcactgagatgaagccggttaatacga 4919
Db 4062 AGCCACCCAGGCGCTCACTGGGAAGTGACTTCACTGAGTAAAGCCGGCTAAATAGGA 4121
Oy 4920 aacaaatctatgtgttggtagaaccttccagaatgagtgagagagctactact 4979
Db 4122 AACAAATTTCTATTGTTGTTTGTACACCTTTTTCAGAGATGGGTAGAGGCTTATCTACT 4181
Oy 4980 aagaagaagactcaaccgctggtgctagaagaatatactggaagaatttccagaatt 5039
Db 4182 AAAAAAGAGACTTCAACCGCTGTGCTAAGAAATACTGAGGAATTTTCCAAAGATT 4241
Oy 5040 ggaataactaaggtataggtcagacaaatgctcagcttgcgttgcagaggtatcag 5099
Db 4242 GGAATACCTAAGTAAATGGGTGACAGCAATGCTCCAGCTTCTGTTGCCAGAGTAAAGTACG 4301
Oy 5100 ggaactggccaagatactggagatgagtggaactgcatctgcatatagagcccaagc 5159
Db 4302 GGACTGGCCAGAGATTTTGGGATTTGATTGAATACTGCAATTTGCTATACAGCCCAAGC 4361
Oy 5160 tcaagacaggtagagagatgaatagaacataaagaagaccttaccataatgacaca 5219
Db 4362 TCAGAGACAGGTAGAGAGATGAATGAACATTAAGAGACCTTACCAATTTGACACA 4421
Oy 5220 gagactgcatatgattgagtgctcctcctgcttgcgtttaggttaggaac 5279
Db 4422 GAGACTGGCATTAATGATTGATGAGTGCCTCTGCCCTTTGCTTTTAAAGGTAGGAAC 4481
Oy 5280 accctggacaagtcttgagctgaacccctatgaattgctctacaggggaaccccccgtg 5339
Db 4482 ACCCTGGACAGTTTGGGCTGACCCCTTAAATTCCTTACGGGGAGACCCCCCGTTG 4541
Oy 5340 gcagaattgctcttgacaatagtgctgaatgctgcttccagccttgccttagg 5399
Db 4542 GCAGAAATTTGCTTTCACATAGTGTGATGCTGCTTTCACAGCTTGTCTTAGG 4601
Oy 5400 ctcaagcagctcagatgagtgagcagcagcagcagcagcagcagcagcagcagc 5459
Db 4602 CTCAAGCCGCTCAGTGGGTGAGGAGCGAGCGTGAACACCTCCGGAGGCTTACTCA 4661
Oy 5460 ggaagagacttgcaatctcacatgctcccaagtggagatcagctatgtagcgc 5519
Db 4662 GAGAGAGACTTGCAGATTTCCACATGCTTCCAAAGTTGAGATTCAGTATTTAGAGCC 4721
Oy 5520 caccgtgcaagaaacctcgagactcggtagaagggaacttactcgttacttgaacaca 5579
Db 4722 CACCGTGCAGGAACCTCGAGACTCGTGAAGGAGCTTATCTGCTTGTGACACA 4781
Oy 5580 ccaagcgtgtagaagtggaagaaatcccaactgtagatccatccagcttaagcg 5639
Db 4782 CCAACGCTGTGAAGTCAAG----- 4803
Oy 5640 ggcacacctccgactcgggtgtagagcgaagaagactggaatcccttaagctgc 5699
Db 4804 ----- 4822
Oy 5700 ctccatgcctgcttcttactcttaacaaatcccaagccagcagtagtaaacgctat 5759
Db 4823 CTCATGCGGTGCTGCTTCTTACTCTGCAATTAACCTCTCAAGTTAATGTAAGCAAGCTTGT 4882
Oy 5760 agacagctggaaccccatagacacttacccttactgctgctgctgctgctgctg 5819
Db 4883 GGAACGCGGCAACTCCATTAACCTTATCTCAGCTGCTGCTGCTGCTGCTGCTGCTG 4942
Oy 5820 gggctgcaatgtaataagacactgaggtgctgctccttagagcagcctgctgagc 5879
Db 4943 AGGTATTAATTAATTAACAGCACTTAAGGAGAGCTGCTGCTGCTGCTGCTGCTG 5002
Oy 5880 gcaatctgctcagctgtagtaacccgctgta-----aaagcaacctcccaact 5933
Db 5003 ATATGCTCGCTCGATAGATATTCCTGCTCAATGACGAGCCACCCCGCATGT 5062
Oy 5934 agtcgtagttagtggttcttcttctgctcagcagc-----caagagaagaagaatgag 5990
Db 5063 ACTCGGCTTACGGGTTTACCTTTCAGGAGAGCCCAATTAATGAAGAAATTTGTGG 5122
Oy 5991 gggctcggggaatcctctgtagagatgagctgctgctcactcccaagctgagagctg 6050
Db 5123 AAATCGTCAGAGATTTCTTTCGAAGCAATGAGACTGCATAACTTAAATGATGGCAATG 5182
Oy 6051 gaaatggcagatctctccaggaacccgggttaaatctcctcttgcataatcc----- 6102
Db 5183 GAATGCGCAGCTCTCAGCAAGACAGAGTAAGTACTCTTTGTTAAATCTTACAG 5242
Oy 6103 ----- 59ccggcgaatgaacaa 6121
Db 5243 TTATTAATCAATTAATTAATGCGCATGGAGATGGAAGAAATGGAACAGCGGTACAAA 5302
Oy 6122 tgaataactataataagataagctgctcccatcagacttagatatacctaagataa 6181
Db 5303 AGATG-TAGCAATTAAGCAATTAAGCTGATTCGTTAGCCATGATTTACTTAATAATA 5361
Oy 6182 gttcactg--aaaggaacacaggaataatctcaaaatgagataatgtagagctgg 6239
Db 5362 GTTCACTGAATAAGAAAGAAACAAAGAAATTTCAAAAGTGGGTAATGCTTGGG 5421
Oy 6240 gaatagtttttat-----tatagcgggggagcagcagcttcaacatcgctccta 6295
Db 5422 GAATAGTACTATAGAGGCTTGGGAGAAAGAAAGAGATCTGTTCTGACTATTTCCCTCA 5481

```

QY 6296 ggatagagacggggaacaacccccctgtggaatgagacccgataaagtactggtcagc 6355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5482 GAATAGAAACCTCAGATGGAACCTCCGGTTGCTATGAGACCAATTAAGGTTTGGCCGAC 5541
QY 6356 aaggagcccccgccctgagcaaccgcaataactgcccgtcccaatlaaactgcgtc 6415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5542 AAGGACCTCCAAATCCAAAG-----ACAGA 5565
QY 6416 ggcctgacataacacagccgctagcaacagtlacacatgagttgattctccacaacgc 6475
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5566 GGCCATTCCTCAACCCCTCTGATTAACAATACACCTCTGGATCACTCCCACTGA----- 5620
QY 6476 ctagaacatccccaggtgtctcgttlaagacagacagagactcttcagtlccagc 6535
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5621 -----GCCATCACTCACTATTAAACAGGGGGCAACTTTTAGCTCATTCACAG 5670
QY 6536 gagcttccagacataccatccacccgactgacactctctctgtgtgtgtgtc 6595
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5671 GAGCTTTTCAGCTCTTAACCTCAACGACTCCAGAGGCTACCTCTTCTTGTGGCTTGGCT 5730
QY 6536 tatcccaaggccctctattatagaggatgctcctaagaagaataatcatgtgcca 6655
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5731 TAGCTTCGGGCCACCTTACATGAGGAATGCTAGAGGAGGAATTCATGACAA 5790
QY 6656 aagacatagaatcatatgataatgaggtgctccgaataagcttaaccctcagtgat 6715
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5791 AGGAACATAGAGACCAATGTACATGAGGATGCCCAAAATACCTTAACCTTACTGAGGTT 5850
QY 6716 ccgggaaggagacatgactaagaaagctcccccatcccaacaacactgtctatgla 6775
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5851 CTGGAAAGGACCTGCTAGGAGATGGATGCTCCCATCCCAACACTTGTGTAAACACA 5910
QY 6776 ctgtgttattgagcaagcccaagaataatcagatttgaactcgtgtatatacagtggt 6835
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5911 CTGAAGCCTTAATGAGAACCTCTGAGACTCAATCTGTGACTGCTGTATGACAGTGT 5970
QY 6836 gggacatgaactcgtgttgaacccccgtgttcccaactcagtlcttaacaaatccaa 6895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5971 GGGCATGTAAATACGTGATTAACCCCTGTGTTCCACCTGTTTCAACCAAACTAAAG 6030
QY 6896 attctgtgatgtgtccaaatgctccccaggtgtactaacatcccgagaagatgtctc 6955
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6031 ACTTTGGCTTATGTGCTCAAAATGTGCCCGGCTGACTACTACTCCGAAACCACTCC 6090
QY 6956 ttgataatgactatcgtatataacgacccaagaagaacccgtatcccttaacctag 7015
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6091 TTGATGATATGACTATATGATATATGCGCCAAAGAGAGCCCATATCCCTGACACTAG 6150
QY 7016 ctgtaatgtcgtatgaagacggcgttggtcgtlaagaaacagggacagctgcccgtga 7075
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6151 CTGTAATGCTCGATTTGGAGTGGCTGCGAGCGTGGGAACAGGAACGGCTGCCATATCA 6210
QY 7076 caggacccacagctagtagaagaagagctgtgtgactacatgcccagacagaaatc 7135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6211 CAGAGCCGCAACAGCTGGAGAAAGGACTTATTAACCTCAACCTCAATTCGATTGTAACGGAATGC 6270
QY 7136 tccagacctlaaaggagctgtttagcaacctagaagaagctccctactcttgcctgaag 7195
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6271 TCCAAAGCCCTTGAATAATCTGTCTAGTAACCTGAGAGAAATCCCTAACCCTCTATCTGAAG 6330
QY 7196 tggttctacagaaacggagagatagatcgtgttcttaagaagaagtggtgtatgtg 7255
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6331 TGGTTTACACAGAACAGAAAGGGGCTTAGATCTGTATTTCTAAAGAGAGAGGGGTATGTG 6390
QY 7256 cagccttaaaagaagaatgtgtctctatgtatagatcactcagaagacatagagactcca 7315
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6391 TAGCTTAAAGAGAGAAATGCTGCTTCTATGATGATCACTAGAGGCCATAGAGACTCCA 6450
QY 7316 tgaacaagcttagaaaaaagttagagagcgtcgaagaagaagaagagcgcgaacaggt 7375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6451 TGAGCAAGCTTAGAGAAAGGTTTAAAGAGGGGTGCAAGGAAGAGAGGCTGACAGGGGT 6510

```

```

QY 7376 gtttgaagatgtgttcaacaggtcctcttgatgacacccctgcttctgtctcagc 7435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6511 GGTTTGAAGATGTTTAAACAGGTCCTCTTGATGACACCTCTCTTCTGCTGACGG 6570
QY 7436 ggccttagtagtctgtctcctgttacttaagtgtggccttgcttaataatagtttg 7495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6571 GGCCCTAGTAGAGTCTCTCTCTGATTAACAGTTGGGCTGCTGTTAATTAATAGTTTG 6630
QY 7496 ttgcttltttagagaaacaggtgagtgcaatccaatcatagttacttagcaacagtlac 7555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6631 TTGCTTTGTTAGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6690
QY 7556 aagcctctcagcgaaggaacatcactcactcagcctccctccagtlcttaagatgaact 7615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6691 AAGGCTCTTGAGCAGAGAGAAATGACCTGATGACCTTCCAGTTCTTAAGATTAGACT 6750
QY 7616 attacaagaacaagaatgaggaaatgaagaatgaagaatgaagaatgaagaatgaaga 7675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6751 ATTACAAAGACAAAGATGGGGAATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 6810
QY 7676 ccaagaaatgaataaaagctcctaagtgcccccgaatlaagacccctgctgtgtcag 7735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6811 CCAGAAATTAATTAATAAGCTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 6870
QY 7736 aaatagtagaaggtcacactcctatgtgtccagggccctgctatccctgagtaag 7795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6871 AAATAGTAAAGTACACTTCTCTATGTTCCAGGGCCCTGCTATCTGCTGCTAATAG 6930
QY 7796 ataacaggaatgagttgactaactcgttactcgtgacttctgaataactgactgacacat 7855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6931 ATAAACAGAAATGAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6990
QY 7856 aagaatltgattacacattgacagccctagtgacatcactacacacacacacacacac 7915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6991 AGAAGAAATTAATTAACATTTGACAGCCCTAGTACCTATCTACATGCAATCTGTACCTC 7050
QY 7916 tgcacaggaagccacagcagatgtgcgaactcccgagactatttlaaatagtatgtgtccag 7975
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7051 TGCCAGAGAGCCACAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 7110
QY 7976 agcggggcctcgtatatttlaaatagtatgtgtccatgtgaagcggcctcgtatatt 8035
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7111 AGCGGGGCTGATATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 7170
QY 8036 aaatgattgtgtgtgacacagcctgtgtgtgaaccccaataaagctgtccagat 8095
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7171 AAATGATGTTGTTTGAGACGACAGGCTTTGTTGTAACCCCAATTAAGCTGTCCGATT 7230
QY 8096 ccgactcggggccgacagctcctacccctgctgtgtgtgaagcgtgtggcccaagcgg 8155
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7231 CCGCACTCGGGGCGGAGTCTCTACCCCTGCTGGTGTAGACACTGTGGGCCCAAGCGG 7290
QY 8156 ctgtgaataaaatcctctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 8198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7291 CTTGGAATAAAATCCTTGTGCTGATCAATCAAAAAA 7333

```

RESULT 3
 US-08-766-528-1
 ; Sequence 1, Application US/08766528
 ; Patent No. 6190861
 ; GENERAL INFORMATION:
 ; APPLICANT: Jay A. Fishman
 ; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
 ; TITLE OF INVENTION: AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 74
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 60 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match 60.7%; Score 4986; DB 4; Length 8060;
Best Local Similarity 95.28; Pred. No. 0;
Matches 5276; Conservative 0; Mismatches 245; Indels 23; Gaps 12;

QY 1 gtgtgtacagctgtgagcccaagcgccttgagataaaatcccttgcgttgcac 60
DB 2530 GTGCTGATGACTGTGGCCCGACGCGCTTGATATAAATCTCTTGCTTGATC 2589
QY 61 aagacgcctctcgtgagtgat-tgggtgtgcctctccagcccgagaggggga 119
DB 2590 AAGACCGCTCTCTGATGATTAAGGGGAGATCCCTTTTCCGAGCTGGAGG- 2642
QY 120 ttgtcttctacgcttcatattgtgtgctgagcgaggaatcccgagacacccct 179
DB 2643 --TTCTTTTCTGCTTACATTTGGGGGCTGTCGGGATCTGTGCGGCCACCTTA 2700
QY 180 acacccagagacgacttgagagta-aaggaatcccttggaaacgtgtg--tgttc 236
DB 2701 ACACCCGAGAACCGACTTGAGGTAATAAAGATCCTTTTAAACGATGATGATACC 2760
QY 237 ggcgggctctctgttcgagtgctgttctcgatgagcgcgcttccggtttgcaatg 296
DB 2761 GCGCGGCTCTCTGTGAGTGTCTGTTTCAGTGTGCGCGCTTTCGGTTTCACACTG 2820
QY 297 tccctcagacgtaagagcagcagcagcagcagcagcagcagcagcagcagcagcag 356
DB 2821 TCTCTCAGGCGCTAAGGCGCTGGGGGAGTGTGATCAGACGAGCGTGAAGAGATCAG 2880
QY 357 gctgcacccctgggggagcccgaggagtgaggagagcagggagcgctgtgtgtctcc 416
DB 2881 GCTGTCGCTCGGGGAGCGCCCGAGAGTGAAGAGACCGAGGAGCGCTGTGTGTCTCC 2940
QY 417 tactgtgagtaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 476
DB 2941 TACTGTGGTGAAGAGCCGAATCTGTGCTGAAGCCAAAGCTTCCCGCTCCGAGACG 3000
QY 477 tccgactcttgcctgtcttgagagagcagcagcagcagcagcagcagcagcagcagcag 536
DB 3001 TCCGACCTTTTGGCTGCTGTGGAATACGTGAAGGAGTCAAGTGTGCTGTGATCTGTG 3060
QY 537 gtctcgtcgtgtgtcttctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 596
DB 3061 GTTCTGTTTGT 3120
QY 597 acagtaacaccccttagttgactcagacatgactgagtaagttaagccagggtc 656

DB 3121 ACAGTACAGACCCCTCTAGTTGACTCGACACTTGAGAGTAAATCCAGGCT 3180
QY 657 cataattgtcagtttaagtttaagagagccttgcagcacttctgtccctcctatgg 716
DB 3181 CATTAATTTGTCAAGTTCAAGTTAAGAGGACCTTGGCACACTTCTGTCTCAATAGG 3240
QY 717 ccaacattgattgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 776
DB 3241 CCGACATTTGATTTGATGAGGCGCATCAGAGGAGCCTTAATTTGAGATTTCTGTGCT 3300
QY 777 gtttaagcaatatttctgaactgagcccgctctatcctgactcagagagccctatc 836
DB 3301 GTTAAAGCAATTAATTTTCAAGCTGACCGGCTCTCAATCCGATCAGAGGCGCTATATC 3360
QY 837 cttagtgagcaagatttgacagaaatccctcagcagcagcagcagcagcagcagcagcagcag 896
DB 3361 CTTACGTGCAAGATTGTGCAAGAGATCTCGCGCATGAGTTAAACATGCTGTAATAG 3420
QY 897 ccaagaaagcaggtcccgaaatcctgtgctcttgagagaaaaaacaacactcggcgaa 956
DB 3421 CCAAGAAAGCAGGTCCCGAATTTGCTGTGAGAGAAAAAACAACTCGGCTGAA 3480
QY 957 aagtcgagcctcttctcgtatctaccgagatcgagagcgcgagcgttgcgga 1016
DB 3481 AAAGTCAAGCCCTTC-TCCTATATATTAACCCGAGATGAGAACCAACGCTTGGCGGGA 3539
QY 1017 acccaactgttcccccacccctatccagacagagtgctgtgagggga-cctctgc 1075
DB 3540 ACCCCAAATGTTTCCCGACCCCTTATCTGACAGAGGTGCGGAGGAGGAGCCTTTGTC 3599
QY 1076 cctctctgagctcgggtgtgagagacgtctgcggagactcggagcggagagagcgc 1135
DB 3600 CCTCTCTGAGCTCCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3659
QY 1136 caaccggag 1195
DB 3660 CACCCGAGCGGAG 3719
QY 1196 gccagggggcgaatttgagccctccagatattgagcccttctctcagatcctataaa 1255
DB 3720 ACCGGGGGGCAATTTGACAGCCCTCCAGATTTAGGCCCTTTTCTGACAGATCTATAPA 3779
QY 1256 ttggaactaacatccccccttctcggagagatcccaagcctcaggggtgtgtga 1315
DB 3780 TTGGAATACTAACCATCCCTTTTCGAGAGATCCCAACCCCTCACGGGTTGTGTGA 3839
QY 1316 gtccctatgtctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1375
DB 3840 GTCCCTTATGTCTCTCAGCAGCTACTTGGGATGATTTCAACAGCTGCTCAACACT 3899
QY 1376 ctccacaacgag 1435
DB 3900 CTTCACAAACGAG 3959
QY 1436 cgaaggggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1495
DB 3960 CGAGGGGCGAGCCAGCGGTTGCAAAATAGATTCAGATGGATTTCCCTTAACCTGCGCC 4019
QY 1496 cagttgagactaacacagcctgaagtagagagagccttgaataatcagcagcagcagcag 1555
DB 4020 CGGTTGGAGCTAACACAGGCTGAAGGTAGGAGACTTGAATAATCTATCGCAGAGCTCT 4079
QY 1556 ggttgagggtctcgggggagcctcaagacggcccaatattgtctaaagtaagagagc 1615
DB 4080 GGTGGGGGCTCTCGGGGGGCTCAAGAGCGCCACTTAATTTGGCTTAAGTAAGTAAGT 4139
QY 1616 gattcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1675
DB 4140 GATGCAAGGAGCGAATGAACCCCTCTGTTTCTTGAAGGCTTTTGAAGCCTTCAG 4199
QY 1676 gcggttcaaccctttagatctactcagagggcccaagagcctcagtgccctggcctt 1735


```

QY 3895 atcaaacctgctctacagagagagtccttgcctccacacagccgctcacaacctg 3954
    |||
Db 6418 ATCAAAAGCCTGCTTCTACAGAGAGGCTCAAGTTCGCTCCACCAACCCCTCTAACCCCTG 6477
QY 3955 ccaactctctgctcgaagaactlgaacagtgatca tga tggccatcaactatga 4014
    |||
Db 6478 CCACCTCTCTGCTGAGAGAGACGTGATGACACAGTGCATGATGCTGCACTCACTATTTGA 6537
QY 4015 ttgagagactggtgtccgcaagaccttacagacataccgctgacgtggagagtgctaa 4074
    |||
Db 6538 TTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAAGTGTAA 6597
QY 4075 cctgttcaactgacgaagaacgactatgtgtgagagtaagagatgctggtggcgag 4134
    |||
Db 6598 CCTGCTTACTGACGGAAGACAGCTATGTGGTGGAGAGTAAGAGATGGCTGGGGCGGGCGG 6657
QY 4135 tgggtgacggaacccgcaagatctgggcccagcagctgcccgaagaaactcagcgcaaa 4194
    |||
Db 6658 TGGTGAAGGAGCCGCAACGATCTGGGCGCAGACCTGCGGGAGAGAACTTACAGCAAAA 6717
QY 4195 aggtctgagctatgtgctcctacagaaacttgcggtggtgcggaaggaataatccataa 4254
    |||
Db 6718 AGGCTGACCTCATGCGCTTCACGCAAGCTTGGCGGTGGCCGAAGGAGAAATCCATAAACA 6777
QY 4255 tttaacgacagcagatgctccttgcgactgcacacgtacacggtggccatcatalaac 4314
    |||
Db 6778 TTTATACGAGACAGCATATCCCTTTGGAGTGCACACGTACATGAGGGCCATCATATAAC 6837
QY 4315 aaaggggtgtcttaacctcagcagggaggaataaagaacaagagaaattcctaagcc 4374
    |||
Db 6838 AAAGGGGTTGCTTACCTACGACAGGAGGAGAAATMAAGAACAAAGAGAAATTTCTAAGCC 6897
QY 4375 tattagaagcttatttgcacaaagagcttatataacatgtgcctggagacataga 4434
    |||
Db 6898 TATTAGAAGCGCTACATTTACCAAAAAGGCTAGCTATTATATACCTGTCTGTGACATAGA 6957
QY 4435 aagccaaagatcatalatcagaggaacagatgctgacgggtgtgccaagcagcag 4494
    |||
Db 6958 AAGCTAAAGATCTCATATCCAGAGAAACAGATGGCTGACCGGGTGGCCAAAGCAGCAG 7017
QY 4495 cccaggtcgttaaccttgcctataagaacgcgccccaaagcccccaagacagcag 4554
    |||
Db 7018 CCCAAGGTTGTTAACCTTGTGCTTAATAAGAAATGCCAAAGCCCCCAAGACCCAGACAGC 7077
QY 4555 agtaacacctgaagactgcaagaagataaagaatagacagcttctcctgaagctcgg 4614
    |||
Db 7078 AGTACACCTTGAAGACAGTGGCAAGAGATAAAGATATGACCA-TTCTCTGAGACTCCGG 7136
QY 4615 agggagactgctatacctcatalatggaagaaatcctgccccacaagaaggttagaat 4674
    |||
Db 7137 AAGGAGCTGCTATACCTCAATGAGGAGAAATCTCTGCCCAAAAGAAAGCGTTAGAAAT 7196
QY 4675 atgtccaac-agatatacgtcttaacccaactagaactaaaccccgagcagtgctc 4733
    |||
Db 7197 ATGTCAACAAGATACATGCTTAACCCACTAGAACTTAACACCTGACACCAATGTGTC 7256
QY 4734 agaacatccccctatcatgctctgaggtacagagatggtgagctggtgtgtaaacat 4793
    |||
Db 7257 AGAACATCCCTTATCATGTTCTGAGGCTACCAAGAGAGTGGCTGACTCGTGTGTAACAT 7316
QY 4794 tgtgtccctgacagctggttaatgtataccttccagataactcaggaagaagacta 4853
    |||
Db 7317 TGTGTGCGCTGCAAGCTGTGTAATGCTTAATCTTCCAAATGCTCCAGGAGAGAGACTA 7376
QY 4854 aggggaagccacccagcgtcactggaagtggaactggaactgaggtgaagccggctaaa 4913
    |||
Db 7377 AGGGGAAGCCACCCAGGCGCTCAGTGGGAGAGTGGACTTCACTGAGGTAAACCCGCTAAA 7436
QY 4914 taaggaaacaatactatgtgttctgtagaacctttagaagatgtgtagagctat 4973
    |||
Db 7437 TATGGAACCAATATACCTTATGTTGTTTGTAGACACCTTTTAAAGATGGGTGAGAGCTTAT 7436
QY 4974 cctactaagaagaagacttcaacggtgtgtgtaagaataatacttgaagaaatttttcca 5033

```

```

Db 7497 CCGTAAAGAAAGAGACTTCAACCGGTAGCTAAAAAAATCTGAGAGAAATTTTCCA 7556
QY 5034 agatttgaatacctaaggaataggttgaagcaatgtgtccagcttctgtgccaagta 5093
    |||
Db 7557 AGATTGGAATACCTAAAGGTAATAGGGTACGAATGTGCCAGCTTTGTTGGCCAGGTA 7616
QY 5094 agtcaagagactggccaagatlatctgggattgatttgaactcattgtgcatacagacc 5153
    |||
Db 7617 AGTCAGAGACTGGCCAAAGTATTGGGATTGATTGGAAACTGATTTGTCATACAGACCC 7676
QY 5154 caaagctcaggaacaggttagagagatgaaatagaaacatlaaagaacccctacaattg 5213
    |||
Db 7677 CAAGCTCAGACAGGTAGAGATGATGAATGAACCATTAAGAGAACCTTAACTAAATG 7736
QY 5214 accaagaaactggcatatgatattgagatgctccctcccttctgttctttaggtg 5273
    |||
Db 7737 ACCGCGAAGACTGGCTTATATGATTGGATGATGCTCTCTCCCTTTGTGCTTTTAAAGGTT 7796
QY 5274 aggaacacccctgagacagtttggctgacccctcctatgaatttgcctacagggagacc 5333
    |||
Db 7797 AGGAACACCCCTGAGACAGTTTGGGCTGACCCCTATGATTTACTGACGGGGACCCCC 7856
QY 5334 ccgttggcagaatgtccttgcacatagtgctgtagtgccttccagccttctc 5393
    |||
Db 7857 CCATTGCTGAAATTTCTCTGATGCTGATGCTGATGCTGCTCTCTCCAGCCTTTGTTG 7916
QY 5394 tctaggtcaagcgctgtagtgggtgaggaagcagcgttgaagcagctccggaggcc 5453
    |||
Db 7917 TCTAGGCTCAAGGCTTGAATGAGTGGTGAACAACAGCCTGAGGCACTCCGGAGGCC 7976
QY 5454 tactc--agagagagacttgcagttccacatgccttccaaattgagatcagctat 5510
    |||
Db 7977 TACTCAGGAGAGAGACTTGGAGATCCACATCGTTTCCAGTGGAGATTCAGTAC 8036
QY 5511 gtttagagccacgctgtaggaac 5534
    |||
Db 8037 GTTAGACGCCACCGTCCAGAGAAC 8060

```

```

RESULT 4
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
; US-09-111-085-3

```

```

Query Match 41.1%; Score 3376.6; DB 3; Length 3482;
Best Local Similarity 99.0%; Pred No. 0;
Matches 3441; Conservative 0; Mismatches 29; Indels 7; Gaps 4;
QY 4718 cctgacagcagttgtcagaacatccctatcatgtctcgaagctacagagtggtcga 4777
    |||
Db 6 cctgacagcagttgtcagaacatccctatcatgtctcgaagctacagagtggtcga 65
QY 4778 cctggtgttcaaacattgtgtccctgacagctggttgaatgtcattccatcagaataac 4837
    |||

```

Dh 66 ctcgtgtgtcaaacatgtgtgtccctgtccagctgtgttaagttaacttccacagataacc 125
Qy 4838 tccagaagaagagactaaagggaagccaccagcgctcaactgtggaagtgagacttcacga 4897
Dh 126 tccaggaagaagagactaaagggaagccaccagcgctcaactgtggaagtgagacttcacga 185
Qy 4898 ggtaaagccggtctaaatacgtgaataaataatctatgtgttttttgaacacctttcag 4957
Dh 186 ggtaaagccggtctaaatacgtgaataaataatctatgtgttttttgaacacctttcag 245
Qy 4958 atggttaagagcttactctactaagaagaagactccaacgctgtgtgtctaaagaataact 5017
Dh 246 atggttaagagcttactctactaagaagaagactccaacgctgtgtgtctaaagaataact 305
Qy 5018 ggaagaaatttttccaagattgtgaatacctaagtaagtggtgcagaacatgttcagc 5077
Dh 306 ggaagaaatttttccgagatttgggaatttccaaagtaacgtggtgcagaacatgttcagc 365
Qy 5078 ttctgtgtccaggttaagtcagggactgtgccaagataattgtgggaattgattggaacatgca 5137
Dh 366 ttctgtgtccaggttaagtcagggactgtgccaagataattgtgggaattgattggaacatgca 425
Qy 5138 ttgtgcatcacagaccccccaagctcagaacagtgtagagaagtgatgaatacctaaga 5197
Dh 426 ttgtgcatcacagaccccccaagctcagaacagtgtagagaagtgatgaatacctaaga 485
Qy 5198 gaccttaaccaatgtgaacacagagactgtgcatatgaattgattgtgtctccctgcacct 5257
Dh 486 gaccttaaccaatgtgaacacagagactgtgcatatgaattgattgtgtctccctgcacct 545
Qy 5258 tgtgctttttaggggtgaagaaacacccctgtgacagtttgggtctgacccccctatgaatgtct 5317
Dh 546 tgtgctttttaggggtgaagaaacacccctgtgacagtttgggtctgacccccctatgaatgtct 605
Qy 5318 ctaagggggagaccccccccgctgttgagagaataatgtgctttgtcacatagtgtctatgtgtct 5377
Dh 606 ctaagggggagaccccccccgctgttgagagaataatgtgctttgtcacatagtgtctatgtgtct 665
Qy 5378 ttcccaagccttgttctctctagagcagaagcgctcgaattgtgtgagagagagagtgtagaa 5437
Dh 666 ttcccaagccttgttctctctagagcagaagcgctcgaattgtgtgagagagagagtgtagaa 725
Qy 5438 gcaagctccgggaagcctactcaag -- aggaagactgtgcaagttccacatcgcttccaaat 5494
Dh 726 gcaagctccgggaagcctactcaagagagactgtgcaagttccacatcgcttccaaat 785
Qy 5495 tggagatttcagctatgttgaacgcacacgctgtgacaggaacactcgaagctcgtgtggaaagg 5554
Dh 786 tggagatttcagctatgttgaacgcacacgctgtgacaggaacactcgaagctcgtgtggaaagg 845
Qy 5555 acctatctcgtacttcttgacacacacacacagcgctgtgaaagtcgaaggaattccccaactg 5614
Dh 846 acctatctcgtacttcttgacacacacacacagcgctgtgaaagtcgaaggaattccccaactg 905
Qy 5615 gatccatgatccacagcttaagccggcgccacactcccgactcggggtgtgagagccggaataa 5674
Dh 906 gatccatgatccacagcttaagccggcgccacactcccgactcggggtgtgagagccggaataa 965
Qy 5675 gactgagaatcccttaagcttcgctccacatcgctgtgttcctactactaataactc 5734
Dh 966 gactgagaatcccttaagcttcgctccacatcgctgtgttcctactactaataactc 1025
Qy 5735 cccagagccagtagtaaaagccttataagacagctcgaaccccccatagaaccttataccctta 5794
Dh 1026 cccagagccagtagtaaaagccttataagacagctcgaaccccccatagaaccttataccctta 1085
Qy 5795 cctgtgctgatatgtgacccctgataaggtgttcaactgttaaatagaactcgaagtggtgtgtc 5854
Dh 1086 cctgtgctgatatgtgacccctgataaggtgttcaactgttaaatagaactcgaagtggtgtgtc 1145
Qy 5855 ctaagagccactgtgtgctggaactgtatcttgcctcgaatgtgataaccccgctgtta 5914
Dh 1146 ctaagagccactgtgtgctggaactgtatcttgcctcgaatgtgataaccccgctgtta 1205

Qy 5915 aaagcacactcccaactagtcctagtagttatgttctatgtctgcacagacagaga 5974
Dh 1206 aaagcacactcccaactagtcctagtagttatgttctatgtctgcacagacagaga 1265
Qy 5975 aagagaatactgtgtgggtgtctgtgggaacacctctgttaaggaatgtgagctgtcaact 6034
Dh 1266 aagagaatactgtgtgggtgtctgtgggaacacctctgttaaggaatgtgagctgtcaact 1325
Qy 6035 ccaacgtatgagacctggaatgtgcccgaatctctctccaggaacgggttaaaattctctgt 6094
Dh 1326 ccaacgtatgagacctggaatgtgcccgaatctctctccaggaacgggttaaaattctctgt 1385
Qy 6095 tcaatttccggcccggtggaatgacaataatgtatgaactatataagaatagagtgcttccc 6154
Dh 1386 tcaatttccggcccggtggaatgacaataatgtatgaactatataagaatagagtgcttccc 1445
Qy 6155 catcagacttagatatactaaagataagtttcaagt -- aaaggaataacaggaataatctc 6212
Dh 1446 catcagacttagatatactaaagataagtttcaagtataaggaataacaggaataatctc 1505
Qy 6213 aaaagtgtataaattgttatgagctgtgggaatgttttctat - tatatgctgggtgagagcag 6271
Dh 1506 aaaagtgtataaattgttatgagctgtgggaatgttttctatataatgtgctgggtgagagcag 1565
Qy 6272 ggtccactttaaccatttgccttagagtagagacgggtgacagaaacccctgtgtgcaatgtg 6331
Dh 1566 ggtccactttaaccatttgccttagagtagagacgggtgacagaaacccctgtgtgcaatgtg 1625
Qy 6332 gaccgcataaagactgtgctggaacagggcccccgccctgtgagccacgcataactgtc 6391
Dh 1626 gaccgcataaagactgtgctggaacagggcccccgccctgtgagccacgcataactgtc 1685
Qy 6392 cgtgtccccaattaaacctcgtgtgagctgtgacataacacagccgcttagcaacagtaacca 6451
Dh 1686 cgtgtccccaattaaacctcgtgtgagctgtgacataacacagccgcttagcaacagtaacca 1745
Qy 6452 ctgattgtattccttaccacaacgctcagagaaccccccaagtgcttccgtttaagaacagac 6511
Dh 1746 ctgattgtattccttaccacaacgctcagagaaccccccaagtgcttccgtttaagaacagac 1805
Qy 6512 agagacttctcagcttccacacagggagcttccaaagcacttcaactccacagaccttagt 6571
Dh 1806 agagacttctcagcttccacacagggagcttccaaagcacttcaactccacagaccttagt 1865
Qy 6572 ccaacttctcgtgtgtgtgttctatccctcaggggcttcccttaattatgagggatgtgcta 6631
Dh 1866 ccaacttctcgtgtgtgtgttctatccctcaggggcttcccttaattatgagggatgtgcta 1925
Qy 6632 aagaagaagaattcaatgtgacccaagaagcatagaataatgtatcatgtgaggttccgaa 6691
Dh 1926 aagaagaagaattcaatgtgacccaagaagcatagaataatgtatcatgtgaggttccgaa 1985
Qy 6692 ataagcttaaccctcacttgaagtttccgggaagggagacatgtca taagaaagcttccocat 6751
Dh 1986 ataagcttaaccctcacttgaagtttccgggaagggagacatgtca taagaaagcttccocat 2045
Qy 6752 ccacacaaacacttgtctatagctagtgtgttataagagagcctcagaataatcagttat 6811
Dh 2046 ccacacaaacacttgtctatagctagtgtgttataagagagcctcagaataatcagttat 2105
Qy 6812 tagtactgttataacaggtgtgtggaatgtgcaatcgtggttaacccccctgtgttcca 6871
Dh 2106 tagtactgttataacaggtgtgtggaatgtgcaatcgtggttaacccccctgtgttcca 2165
Qy 6872 cctcagcttccaacaaatccaaagatltctgtgtcatgtgtccaaatcgtlcccccagtggt 6931
Dh 2166 cctcagcttccaacaaatccaaagatltctgtgtcatgtgtccaaatcgtlcccccagtggt 2225
Qy 6932 actacatccttaggaagtggtccttgtatgaatagactatcgtgtataacggaacaaa 6991
Dh 2226 actacatccttaggaagtggtccttgtatgaatagactatcgtgtataacggaacaaa 2285

QY 6992 gagaaccgtaaccctaccctagctgtaatgctgagattaggagcgccgttgcgtag 7051
|||||
Db 2286 gagaaccgtaaccctaccctagctgtaatgctgagattaggagcgccgttgcgtag 2245
QY 7052 gaaagagagagcgctgctgctgatacagagacacagcgctgagagaaagcgcttgcgtagc 7111
|||||
Db 2346 gaaagagagagcgctgctgctgatacagagacacagcgctgagagaaagcgcttgcgtagc 2405
QY 7112 taatgctgagccatgacagaaatctccgagccctaaagagctgctgtagaaccctagaag 7171
|||||
Db 2406 taatgctgagccatgacagaaatctccgagccctaaagagctgctgtagaaccctagaag 2465
QY 7172 agtcccgtaactcttctgtaagtggtcttacaagacggagggagctgtagctgctg 7231
|||||
Db 2466 agtcccgtaactcttctgtaagtggtcttacaagacggagggagctgtagctgctg 2525
QY 7232 ttccaagaagagtggtgttctgtaagcttacaagaaagagtggtgcttctatgtagatc 7291
|||||
Db 2526 ttccaagaagagtggtgttctgtaagcttacaagaaagagtggtgcttctatgtagatc 2585
QY 7292 actcagagagccatgacagaaatctccgagccctaaagagctgctgtagaaccctagaag 7351
|||||
Db 2586 actcagagagccatgacagaaatctccgagccctaaagagctgctgtagaaccctagaag 2645
QY 7352 gggaaagagagtgctgacagaggggtgttgaaggaagtgcttcaacaggtctccttgatga 7411
|||||
Db 2646 gggaaagagagtgctgacagaggggtgttgaaggaagtgcttcaacaggtctccttgatga 2705
QY 7412 ccaccctgcttctgctctacaggggcccctagctgctgctcctcttcttactaagtg 7471
|||||
Db 2706 ccaccctgcttctgctctacaggggcccctagctgctgctcctcttcttactaagtg 2765
QY 7472 ggccttgccttaataataggttctgctccttctgtaagagacagtgagtgctcagca 7531
|||||
Db 2766 ggccttgccttaataataggttctgctccttctgtaagagacagtgagtgagtgctcagca 2825
QY 7532 tcatgctgacttaggcaacagtagcaagggccttctgagccaaagagaaactgagccttagc 7591
|||||
Db 2826 tcatgctgacttaggcaacagtagcaagggccttctgagccaaagagaaactgagccttagc 2885
QY 7592 ctcccgagcttgaagtagtaacattaaacaagaagaagtgaggatgaataaagagtagaa 7651
|||||
Db 2886 ctcccgagcttgaagtagtaacattaaacaagaagaagtgaggatgaataaagagtagaa 2945
QY 7652 aatgcaacctaaccctcccaagacccaggaagtagtaataaaagccttaaaatgccccgaa 7711
|||||
Db 2946 aatgcaacctaaccctcccaagacccaggaagtagtaataaaagccttaaaatgccccgaa 3005
QY 7712 ttaacagacctgctgctgctgcaagtaaatagtagaaggtcacacctctatgttccag 7771
|||||
Db 3006 ttaacagacctgctgctgctgcaagtaaatagtagaaggtcacacctctatgttccag 3065
QY 7772 gccctgataccctgctgctgcaagtaaatagtagaaggtcacacctctatgttccag 7831
|||||
Db 3066 gccctgataccctgctgctgcaagtaaatagtagaaggtcacacctctatgttccag 3125
QY 7832 ttctgtaanaactgctgctgcaacatagagaatgattatcacatgtaagacccctagtgacc 7891
|||||
Db 3126 ttctgtaanaactgctgctgcaacatagagaatgattatcacatgtaagacccctagtgacc 3185
QY 7892 tatctcaactgcaatctgctgctgcaagggagccacagcagatgagagccctcggagc 7951
|||||
Db 3186 tatctcaactgcaatctgctgctgcaagggagccacagcagatgagagccctcggagc 3245
QY 7952 taatttaaaatgattgctgcaagcgagcgccctcgatattttaaaatgattgctgca 8011
|||||
Db 3246 taatttaaaatgattgctgcaagcgagcgccctcgatattttaaaatgattgctgca 3305
QY 8012 tggagcgcgggctc-tcgataattttaaaatgattgctgcaagcgagccttctgt 8070
|||||
Db 3306 cgagcgcgggctcctcgataattttaaaatgattgctgcaagcgagccttctgt 3365
QY 8071 gaaccataaagctgtcccgatcccgactcgggcgagctcctctacccctgctg 8130
|||||

Db 3366 gaaccataaagctgtcccgatcccgactcgggcgagctcctacccctgctg 3425
QY 8131 ggtgtaagctgtggggcccaagcgcgcttggaataaaatcctctgtcttgatc 8187
|||||
Db 3426 ggtgtaagctgtggggcccaagcgcgcttggaataaaatcctctgtcttgatc 3482
|||||
RESULT 5
US-09-075-272-1
; Sequence 1, Application US/09075272
; Patent No. 6136598
; GENERAL INFORMATION:
; APPLICANT: MILLER, A. DUSTY
; APPLICANT: WOLGAMOT, GREG
; APPLICANT: BONHAM, LYNN
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
; TITLE OF INVENTION: PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/075, 272
; FILING DATE: 08-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,140
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: POOR, Brian W.
; REGISTRATION NUMBER: 32,928
; REFERENCE/DOCKET NUMBER: 14538A-003710
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8655 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-09-075-272-1
Query Match 28.0%; Score 2295.8; DB 3; Length 8655;
Best Local Similarity 60.3%; Pred. No. 0;
Matches 4362; Conservative 0; Mismatches 2662; Indels 209; Gaps 26;
QY 464 cctcggcgccgctgagctcttgcgcctgctggaagagcgagcgagcggtcgtg 523
|||||
Db 601 CCAGCCCATTCATCTGAATCTTTAAGTTGTGTGTGTGACGCGCAAGTCCGCCGGTT 660
|||||
QY 524 tctgatac-tgtgtgttctgctgctgctgtctgtctgtctgtcgtcctgtctacagtc 582
|||||
Db 661 TTGGTTCTTTTGT 716
|||||
QY 583 ttaataatggagacagagtgactacccccctagttgactctgacatgagctgagag 642
|||||
Db 717 TAGAATGGAGCAATCTGTCTCCACTCCCTTTCTTAACCTGGAGCATTTGAAGAGG 776
|||||
QY 643 ttgatacagggctcataatgttgcagtgaggttaaaagagagcttgacagcttct 702
|||||

Dh 5100 CCAAGAGATTGTAAGAAATCCCTGCCAAGATGAGTACTTAAGGTAATGGGTCGG 5159
Qy 5065 acaatgtccagcttcctgttcccaagtaagtcaggagacttgcgaagataatltggagatg 5124
Dh 5160 ACAATGACACAGCCTTGTGTCCAGGTAAGTACAGGATGTGGCCACTCACTTGGGATG 5219
Qy 5125 attgnaactgcatltgtacacagaccaccaagctcaggaacggttagagagatgata 5184
Dh 5220 ATTGAATTAACATTTGTCTTACCGCCCTCAAAAGCTCAGACAGGTAGAGAGATGATA 5279
Qy 5185 gaaccataagagacccttaaccataattgacacagagacttgcgataatgattgagag 5244
Dh 5280 GAACATTAAAGAGACCTTGACTTAATTTGGCCATGGAGACCGGGGGAAGAACTGGGTGG 5339
Qy 5245 ctccctcccttctgttctttagggtagaagaccccttgcagatgttgagctgagcc 5304
Dh 5340 CTCCTCCCTGCTTGGCTCTTCCGAGCCCGAAGACCCCGGAGCTTTTGGGCTCACTC 5399
Qy 5305 cctatgaatgtcttaagggagagaccccccgcttgcagaaatg-----ccttgcac 5358
Dh 5400 CTTTGAAGTCTGTATGGGGGACCTCCCTTAATTAAGATGGTGAACATTGGTTTC 5459
Qy 5359 atagtgatgtgtgtgttccagacctgttctctcagagctcagagcgctcagtgag 5418
Dh 5460 CCGATTCAAGCTGTCTTACCTCTCTTGTCTTATTCATTAAAGGCCCTGAAGTAA 5519
Qy 5419 tgaagcagcagcgttgaagacagctcggagagcctactcagagagagacttgcagatlc 5478
Dh 5520 TTAGAACCCAAATTTGGGACACAGCTTAAGAGAGGCTTACACCCCGAGGACCCGACATAC 5579
Qy 5479 cacatcgcttcccaagtttgaagattcagctctgttgaagcgcacacgctcagaaacctgc 5538
Dh 5580 CCCAGAGATTCCAGAGTCGGGAGACCAAGCTTGTGTGACAGACATGAGAACGGGTAGCCTG 5639
Qy 5539 agactcggtagaagagacctatctcgtacttgaaccacacagcgctgtgaaagtcg 5598
Dh 5640 AACCAAGGTGAAGGAGACCTTATTAAGTACTTAACAACTCTTACGCGACATGAAGTGTG 5699
Qy 5599 aaggaatcccccccgatgataatgataccagcttgaagcggcgcc----- 5644
Dh 5700 ACGGATTTGCTCTGGATCCACAGCTTACACAGTCAAGAGAGGCCCCCAAGTCAAGTGAAG 5759
Qy 5645 -acctccagactcggggtgagagcgaagaaagactgaagatcccccttaagcttgcctcc 5703
Dh 5760 AAACCCACGAAACAACTGGGCTGTGGAAGCCACTGATTAACCTTTAAGCTTGGCTTCG 5819
Qy 5704 atcgccgtgttcttacttactcaaat----- 5729
Dh 5820 GTGCGAGAGAGCCCTCCATCACCTCGACCTTAGGGAACACACACCTCATGCCAGTT 5879
Qy 5730 -----aaactcccgagcgagtagtaaacgccttataagacagc-----tc 5768
Dh 5880 CAACACTCTTGGGAAGTGTCTTAATGAAGAGAGACGTTGTATGGGTAGCCACTGCAGTC 5939
Qy 5769 gaacccecaatagact-----ttatcccttaacctgcgtatattat 5807
Dh 5940 CATCCCTTGGACTTGGGGGCTGATCTCACACCTGACATCTGTAATTAACACAGGA 5999
Qy 5808 tgaacctgaacggtgtcactgttaataagcaactcggaggtgtgtcctcctcagagagactgc 5867
Dh 6000 TCTCCCAATTTGGGACCTTCCGATCATCTGACTTGAATTAACCCACCTCTGAACAAAAG 6059
Qy 5868 gtgagcctgaacgtcatttctgcctcagattgattaaacccgctgttaaaagacacctcc 5927
Dh 6060 TGTGTCCCAACAGGGGTAGGAAGCACTACTGATGTTCGGGGCAGTTTATGAGAGTAT 6119
Qy 5928 caactagtcgttagttagttaggttctactgtcgtccagagcagagagaa-----gagaa 5982
Dh 6120 CTTAGGGGCTGCACAGTTTATGTTTGGCTGTGCAAGGTGAGAAAGAAAGCTGCAACAA 6179
Qy 5983 tactgtgaggttctcgtgaggaactctctgttagagagatgagctgcgtcaccctcaagat 6042
Dh 6180 GAATGTAGAGGGGCACTCACTTTTGGGGTAATGGACATGTGAACCAACAGGGGGA 6239

Qy 6043 ggaagactgaa-----atggcattctctccagagcgggttaaaattctccttgtcaa 6098
Dh 6240 GCTTACTGGAAGCCTTCCCTGATTTGGAGCCTGATCAGGTAAACGTGGTATGTTAT 6299
Qy 6099 ttccggcccggtgcaagtaacaaatgataagtaataagataagagctctcccatc 6158
Dh 6300 GATTAACCAACCAAGAGAGAAAGAAACCCATCAATTAACCTGATTTGGGTGCTCTT 6359
Qy 6159 agacttagattctaaagataagtttcaactgaagaaagacaggaataatc-----aaaa 6216
Dh 6360 AAAAATTAACAGCCGCCACAGACCATGCAAGTAAATACCTGACACCCCTACTCAATAA 6419
Qy 6217 gtgataaatgtatagcttgggaaatgttlttaltatataatgctggggagagaggtcc 6276
Dh 6420 TTACTGAGAAAGGAAACAGAGTCCCTGATGTGGCTTAAAGAAATAGTGGGGTTGG 6479
Qy 6277 acttaacattcgccttagatagagacgggagacacccctgtgcaatgtgagacc 6336
Dh 6480 CGAGTATATATTCATTAAGAGACCCCTGGGTTTATCTTTACATTAAGACTGACATGAGA 6539
Qy 6337 gataaagtactgtctgaacaggggcccccgccccgtgagccacagcataactgtccggtg 6396
Dh 6540 GACCTGGCACTTAATTCATAGAGACCAACAGGTCTTAAGAGAGAGAGAGAGAGAGT 6599
Qy 6397 ccccaatlaacctgcgtcgtgagacataacacagcgctagacagagtlaccactgga 6456
Dh 6600 GCACCGGCTCCCGAGAGTCCAGCGGTCCAGTCCACCACTTACAGGCGCTTACACA 6659
Qy 6457 tgaatcctcaacacagcttagaactcccccgagtggttcctgttgaagcagagagaga 6516
Dh 6660 GTAGGACCTTATTAAGAGACACCTTAAGCTTCCCAACACTCTTAATACAGAAACCT 6719
Qy 6517 ctcttaagctcaacagagagcttccaaagcacaactcaactccagacccctgtgcaact 6576
Dh 6720 CTGTCTAGTGTAGTTCAGGAGGCTTTTGTATTAATGAGCTAATCTAATATGACT 6779
Qy 6577 tctctgtgtgcttgcctatcctcagagcctccttattatagagagatgagctaaaga 6636
Dh 6780 CAATCATGTTGTTATCTATGCTTACACCCCTTATTAAGAGAAATGTCTGAGACT 6639
Qy 6637 agaaatcaatgtgagcacaagagcagatagaatcaatgatacatgaggtgtccgaataag 6696
Dh 6640 AGAATTAACATTAATTAATCAATGATCTC---TCAATGTCTTTGGGAGAGAACAGGAAG 6896
Qy 6697 cttaacctcaatgaagttccgggaagggacatgacatagaaagagctcccccatccac 6756
Dh 6897 TTGACTGTGACAGAGTTCAGAAATGGGCTTGTATGAGTCAAGTGGCCAGAGATAA 6956
Qy 6757 caacaccttgcataagtagctgtgttataagagagccctcagaaataagattatagta 6816
Dh 6957 TGGCACCTCTTACACAGACTCAAAATATCCGACTTAACAAAGGTGGTCAATCTAGTG 7016
Qy 6817 ccttgattaaacaggtgtgtggaatgcaatcgtgttlaacccccctgtgttccacatca 6876
Dh 7017 CCTCCATAGACACAGATATGGGCTTGCATATGAGTCTCACTCTTGTATATATGATGCT 7076
Qy 6877 gtcttcaaccaatcaaaagattctgtgtcattgtlccaatcgtgtcccccgaggttactac 6936
Dh 7077 GTTTTCAATTAAGCTCAAAAGATTTCTGTATTTAGTTAGCTTAATCTCTAGACTCTGTAT 7136
Qy 6937 catcttgaggaaggtgtccttgaatgaatgacatcgtatataacacaaagagaa 6996
Dh 7137 CATGATGATACCTCATTTCTGACAAATTTAACAATCGG---GTCCGCTGAAAAGAGAA 7193
Qy 6997 cccgtatcccttaacctagctgtlaatgtcgtgalt---aggagagcgctgtgtgcttaga 7053
Dh 7194 CCGATTACTTTAAGCTTATGCAAGTCTTTAGAGTTGGAGTGGCCGACGAGTGGGTAGGT 7253
Qy 7054 acaaggacagctgcctgtatcaacagacacagcagcttagagaaagactgtgtgagcta 7113
Dh 7254 ACTGAAACCGGTGCTTATATCCAGACCCCGGATACTTTAGAG-----GAATTA 7301

```

OY 7114 catgagccatgacagaagatctccagccttaagaagatcgtttagcaacctagaagag 7173
Db 7302 GGTACAGCTATGATGACTGATCTCAGAGCTTAGAACACTATTAACCAACTAGAGAA 7361
OY 7174 tcccgactctcttgctgaagtgctctcaagaacggaggagatagatctcgctgtt 7233
Db 7362 TCTTAACTTCTGCTGAGAGTGTACTGCAAAATAGAGAGAGATTAGATTGTATTTT 7421
OY 7234 ctaagaaggttggtttagtctgcacgcttaagaagaatgttctctctctatagatc 7293
Db 7422 CTTAAGAGAGAGACTTTGCTGCTCTCAAGAGAGAAAGTGTGTATATGCGACAC 7481
OY 7294 tcaagagccatcagaagactcatalgaagaacattagaagaagtgagtgagag 7353
Db 7482 TCGGAGAGATCAAGATTTCTATGCGCAACTTAGAGAACGCCCTATATAGCTCAAGA 7541
OY 7354 gaaagagagctgacaggggtgtttagaagatggttcaacagctctctctgagatgac 7413
Db 7542 GAGAGAGAGAGCAAAACAAAGATGTTGAAGAGCTGTTAATTAAGTCCCTGCGTCAAC 7601
OY 7414 accctgcttctgctctgacggggccctagtagtctctctctctctctctctctct 7473
Db 7602 ACTCTCTTTCACATAAGGCGGACCCCTGATATATCTTCTGCTTTGCTTACTTTGGC 7661
OY 7474 ccttgcttaataataggttctgtctgtttagaagaacggtgagtgagtcagatc 7533
Db 7662 CCGGCACTTCTTAAATAGTATAGTACCTTTATTTAGAGAAATTAATGCACTACAGGTT 7721
OY 7534 atgtacttaggcaacagtiacagagcctctg 7566
Db 7722 ATGTTATTAACACAAATATCTAGTCTTCCAG 7754

```

RESULT 6

```

US-08-716-351A-1
; Sequence 1, Application US/08716351A
; Patent No. 6033905
; GENERAL INFORMATION:

```

```

; APPLICANT:
; TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based
; TITLE OF INVENTION: Retroviral Vectors
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/716.351A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03784
; FILING DATE: 06-APR-1994
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 15280-128-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-5043
; TELEFAX: (415) 543-9600
; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 8535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:

```

```

; NAME/KEY: misc.feature
; LOCATION: 1..8535
; OTHER INFORMATION: /standard_name= "GalV SEATO Genome"

```

US-08-716-351A-1

Query Match 27.3%; Score 2244.4; DB 3; Length 8535;
 Best Local Similarity 59.1%; Pred. No. 0;
 Matches 4366; Conservative 0; Mismatches 2811; Indels 215; Gaps 22.

```

OY 332 agcgaagctgctgaagagatcaacagctgcacacctgaggagagcccgaggttgagga 391
Db 756 AACAGAGGTGTCGGGGCTCACCGCTGTAAATCCAGGGAGAGAGCTCCAGATCAAGGGGA 815
OY 392 gaacgaagagcctgtgtgtctctctctctctctctctctctctctctctctctctct 451
Db 816 GGACCAAGGAGAGCTGTGTGACCCCTCG-----GTAAAGGCTGTGTATACCG 864
OY 452 gcaagaagctccctcccgccgctcgcactctctctctctctctctctctctctctct 511
Db 865 ATTTGATGCGCCCTGTTGTTGAAGAGCGCTGTGAATCTGATTCMTCTCGGTCCCTCG 924
OY 512 ggttcgctgtgtctgtgatctgttctctctctctctctctctctctctctctctct 571
Db 925 CCGCGCTCTGTGTCTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 984
OY 572 tgcataagtttaataatggaagacagtgactaccccccctctctctctctctctctct 631
Db 985 CTCCTCAATCATCATATATGAGCAAGATTAATTTCTACCCCTATCTCTCTCTTAATCA 1044
OY 632 ttgactgaagttagatccagagctcaatatttctcagttcaggttaagaagagacctg 691
Db 1045 CTGAGAGATGTGAGAACAAAGGGCTCACATCTATCCGTGGAATCAAAAGGGGAATG 1104
OY 692 gcagactctctgtcctcgaatgccaactctgattttagatggccatcaaggagac 751
Db 1105 GCAGACTTTCTGTCTCTCGAGTGGCCCAATCTGCGGCGGCGGCGGCGGCGGCGGCGG 1164
OY 752 cttaattctgaatlatcctctgttgaagcaatcttctcagactgagaccgctc 811
Db 1165 TTTTAATCTCTCTGATTTTTCAGTTAAAGATGTCTTTGAG---GAGAACGGGG 1221
OY 812 tcatctctgagagagccatctctctctctctctctctctctctctctctctctctct 871
Db 1222 ACATCGGACCAAGTTCCATATATCTGTATGAGAGACCTCCCGACGATCCCGAC 1281
OY 872 atgggttlaaacatctgcttaataaacaagaagcaggtcccgaaatctctctctct 931
Db 1282 ATGGGTGCGAGCTCCGCGAAGTCCGTGTGTCTGTATACCGAAGACCATGTCGGG 1341
OY 932 agagaaaaaacacactcgcgcgaaaaagtgcagccctctctctctctctctctct 991
Db 1342 GAGGCCATCAGCTCCCTCCCGACCCCGCATCTACCCGCAACAGACACTTACTCTCC 1401
OY 992 tcgaagagcgcgcgactgtgcggaag--cccaacgcttcccccacccctatccagc 1049
Db 1402 CTCGTACCCACAGCCCGCCGCTTATCCGGGCGACCTCCACCCCTCTGCGCCCTCAGGC 1461
OY 1050 acaagggtctgtgaaggagactctgcccc--tctgtgagctcgggtgtgtgaaggagctc 1108
Db 1462 GATCGAGCCGCGCTCAGGCCAGATGCCGATGATGAGGATCTGAGAGGGCGACCCCTG 1521
OY 1109 tgcgagactcgagag-----ccgagagagcgcaccccgagagcagagatctcgat 1165
Db 1522 GACCAAGGAGTCCGCTCCGCGACATCCAGCACTGCGGCTCGACTCCACATGATGAT 1581
OY 1166 attacgctgcgcacatagccctcccatgcaagcagggggccaattgcagccctcagatg 1225
Db 1582 TTTGCCCTCGAGACCATATAGGACCCCGCGAGGCCAATGGCTGTGCTCTCAATA 1641
OY 1226 ttgaccttctctctgagatctataattggaataacacatccctctctctcga 1285
Db 1642 TTGGCTTTTTCCTCAGACAGATCTTATATGGAATCTAATCATCCCTCTTTTTCGA 1701
OY 1286 ggtcccaacgctcagaggggttggtgaggtccctatgctctctctcagcagactactg 1345

```

Db 1702 AAACCCAGCAGCTCTACGGGGCTCTTGAAGTCTTATGTTCTCCATCAGCCACTG 1761
QY 1346 ggaatatttcaacagctctgctcagacaccltcacacccagagcagagaatctc 1405
Db 1762 GGAGATGTGCAACAGAGCTCTACAGATTCTTTTACCACTGAGAGACGGGAAGATTTCT 1821
QY 1406 gtttagagctagaanaaatgtctcctgggagcgacgagcgagccagctgtcaaaatga 1465
Db 1822 CCTGGAGGCGCCGCAAAATGTCTTGGGGACAAATGGGGCCCTACACAGCTCGAGACCT 1881
QY 1466 gattgaactggaatttcctcctgactgcgcgcgtgttgagctacacacgcgtgaagtag 1525
Db 1882 CATTATATGAGGCTTCCCTCAATGCACTCTACCTGGGATTTACACACCCGAGGTAG 1941
QY 1526 ggaagacttgaanaactatcagcagcgtctgtgtgaggtctccggggcgccctcaagag 1585
Db 1942 GGAGGCTGTCTGTGCTTACCGCCGACTCTAGTGGCAGGTTCTCAAAGGGGCGAGCTCGGG 2001
QY 1586 gcccaactaatttggctaaagtaagagagtgatgcaaggacccgagacgacctccctcgt 1645
Db 2002 TCTTACCAATTTGGCTAAGGTAAAGAGAGGTCTTCAGAGGACCGGACAGACCCCTTGCT 2061
QY 1646 attcttgagaggtctatcagagctcagcggttcaaccttttgaactcactaga 1705
Db 2062 TTTCTTGAACGCTGTGATGAGGCTTATGAGATATACCTCCGTTTATCCCTCTTGTA 2121
QY 1706 ggcacagaagccctcagctgagccctgacttattgggcagtcagctcgtatcagaa 1765
Db 2122 GGCACACAGGCTTCGGGTCCGCAATGGCTTTATGGACAGTCAAGCCCAATATCAAGAA 2181
QY 1766 gaaactcagagacttgaaaggttaccagagagctgagttacgtatcagtagtgaagagc 1825
Db 2182 AAATTTACAGGCTTACAGGGCTCCAGGACTTCTCTTCAAGATTTAGTAAAGAGGC 2241
QY 1826 agagaaggttattaccagagagagacagagagagagagagagagagagagagagagag 1885
Db 2242 AGAAAGGCTTACCATATAGAGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2301
QY 1886 gagaagaagaagagagagagagagctgtatagagcgcaagagagagagagagagagag 1945
Db 2302 GGCAG 2361
QY 1946 ggcgcagctggtt-tgaag 2003
Db 2362 GGCAG 2421
QY 2004 aggtcagagccctagacagctcagagagagagagagagagagagagagagagagag 2063
Db 2422 GGCAG 2470
QY 2064 tgtgctattgttaagaagaagaagagagagagagagagagagagagagagagagagag 2120
Db 2471 TGCCATATCTGTATAG 2530
QY 2121 aagaagacgaaggtcctagctctagaagaagaatagaagagagagagagagagagagag 2180
Db 2531 AG 2581
QY 2181 cccctcccgagccagaggttaacttgaaggtgagagagagagagagagagagagagagag 2240
Db 2582 CCCCCTCCCAACCTAGAGGTACAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2641
QY 2241 gatacagagcagagacatgaatgctcagcaacatltgagaaacttaaaagaaaaa 2300
Db 2642 GACACCGAGCTGACATTTATGATTTGACCAACCCATGGGAAAGTAGGAGGCTCAACAGG 2701
QY 2301 tcttgaggtgaggtgagcag 2360
Db 2702 ACGGTCTGAG 2761
QY 2361 gacttgagaggtgag 2420
Db 2762 AAATTTGACATTAACAAAGTACCCACTCTTCTGTGTATACCGAGTGCCTCTGCTCT 2821
QY 2421 ctctagtagagacttactagaccaaagatgagagagagagagagagagagagagagagag 2480
Db 2822 CTGTTGGCAG 2881
QY 2481 coagaaggtctgtgataaacaacacacacacacacacacacacacacacacacacacac 2540
Db 2882 CCACAGGTAAACATGGGAG 2941
QY 2541 taccagactatctcccaagaaagcctgtatcaagatatacaggtccctggttgaagag 2600
Db 2942 TACCGACTCATGTAAG 2998
QY 2601 ttccccaagccctgggcaagaag 2660
Db 2999 TTCCCACTGTATGGGAG 3058
QY 2661 gttatcaactgaagagcaggtgtacacagatcagtcagacagtagtcccttgaataga 2720
Db 3059 GTAGTAGAGCTAAGATCAGGTGCTTCCAGAGTGTTCGCAATATTCATATGAGCAAA 3118
QY 2721 gaggctcag 2780
Db 3119 GAAGCTCGGGAGAGTATGAG 3178
QY 2781 cctgtccaatcccttggaatactccctgctacaggttgaagagagagagagagagagag 2840
Db 3179 CCTGTGCGGCTCGCCCTGGAGATACCTCTGCTACCTGTATTAAGAGAGAGAGAGAGAGAG 3238
QY 2841 taccagacagtag 2900
Db 3239 TATCGGCACTGTATGAG 3298
QY 2901 gtcggaacccctataactcttgagcgcctcccgctcagagagagagagagagagagag 2960
Db 3299 GTCCCAACCCCTTACATCTTGTGAGTCCCTTCGCTTACGCTATATCTGTATCTAGTC 3358
QY 2961 ttgagactaaagatgctctctctcctcagagagagagagagagagagagagagagagag 3020
Db 3359 TTATGATCTAAGAGATCTCTTTTCTGCTTCAAGCTATATCCCAACAGCCAGGCTGTTC 3418
QY 3021 gcttcgaatgag 3080
Db 3419 GCGTTGAGTGGAG 3478
QY 3081 ccccaaggttcaagaag 3140
Db 3479 CCACAAGGCTTCAAG 3538
QY 3141 aactcagagatcaaac 3200
Db 3539 CCTTTAGGGCCCTCAACCCCAAGGTGCTTACTCTCAATATGTGAGACACTCTTGGTG 3598
QY 3201 ggcgagccaccaaag 3260
Db 3599 GCGGCCCCCAATATGAG 3658
QY 3261 gacctaagctacagagcctctgctagaagaagagagagagagagagagagagagagag 3320
Db 3659 AAGTTGGGTATCCGGTATTCGCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3718
QY 3321 ttggggtatagtttcgag 3380
Db 3719 CTGGGCTACTCTCAAG 3778
QY 3381 gtcagataccgagcccaac 3440
Db 3779 ATGAAATATCCCTGTTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3838
QY 3441 ttctgag 3500
Db 3839 TTCTGAGAGCTGTGATCCCTGGGTTTCTTCCCTGCTGACACCTTGTACCCCTTAACA 3898

Db 2046 GTTGCTGCTACTGCTGTTAGCGGGACAGACAGATAGACAGGAGAGAGCCAAAGAG 2105
QY 2042 gaccccaatcgacaagaaccagctgtgctatgttaagaanaaagacactgggcaagaa 2101
Db 2106 GCCCAATCTCACACGACGACGCTGCTACTGCAAAAGAAAGGAGCATTTGGCTTAAGA 2165
QY 2102 ctgcccagaagaaggaaacaagaagcgaagtccta-----gcttaagaagaga 2152
Db 2166 TTGCCCAAGAGACAGAGAGACCCCGGGACACAGACCCAGGCTTCCTCTGACCTT 2225
QY 2153 taagaatlagggagagcggtgtcgaaocccctcccgagcccaagggttaacttgaagt 2212
Db 2226 AGAGATTAAGGAGGTAGGGTCAAGAGACCCCTCGAACCAGATTAACCTTCACAGT 2285
QY 2213 ggaaggcgacaaccagttgaatctcgtgttgataccggaagcggaagcattcagtgctgtaca 2272
Db 2286 CGGGGGGCAACCCCTCACCTTCTAGTGATACGCGGGCCCAACACTCCGTGTCAGCCCA 2345
QY 2273 accatlaggaactaaagaanaaaatcctgggtgagtgtgccaagggcaagga 2332
Db 2346 AAATCTGAGACCCCTTAAGTGAAGTCTGCTGGCTCAAGGSGCTACTGGAGGAAAGCG 2405
QY 2333 gtaaccatggaactaaccgaagaacgcttgacttggaagtggaaggttaaccactcgt 2392
Db 2406 GTATGCTGACACGACGATCGCGAGTGCACCTAGCCACCGTAAAGGTCAACCTTCTT 2465
QY 2393 tctgtatccctgagtgagtgaccccttctaagttagagacttactaccaaagatgga 2452
Db 2466 CCTCATGTACCAATGATGCTCCCTTCTGCTAGGAAGATTTGCTTAAACCTTAA 2525
QY 2453 agcccaaatctcttga---acaaggaagccagaagtgctgtgataacaacccat 2509
Db 2526 AGCCCAATTTACTTTGAGGATATCAGAGAGCTCAGGTTGTGGACCAATGGGACAGCCCT 2585
QY 2510 cactgtgtgaacctccaatatagaatgaatatacgactatctcccaagtaaac 2569
Db 2586 GCAAGTGTGACCTTAAACATAGAGATGATGCTATGCTACATGAGACCTCAAAAGGGCC 2645
QY 2570 tgaatcaagatatagctcc---tgtgtggaagcagttccccaagcctgggcaagaacgc 2626
Db 2646 AGATGTGCTCTTGAAGTGCACATGCTCTGATTTTTCCTCCAGGCTGGGACAGAACCGG 2705
QY 2627 aggaatgagtttgcaagaagcagttcccccacaggttatcaactgaagccagtgta 2686
Db 2706 GGGCATGGGGCTGGCCGTTGCCCAAGCTCCTCTATCATCTTGAAGCAACCTTAC 2765
QY 2687 accgaatcagtcagacaagtaaccccttgagtagaagcctcgagaagaaatttgccga 2746
Db 2766 CCGCGTGTCCATAAACATACCCCATGTGCACAGAAGCCAGACTGGGGATCAAGCCCA 2825
QY 2747 tgttcaagaatataccaagaagcagctcagttccgtgcaatcccttggaaatactc 2806
Db 2826 CATTCAGAGACTGCTGGATTCAGGGAATTCGTGACCTGTGAGTCCCTGGACACAGCC 2885
QY 2807 ccttctaaccggttagaagccttggaccaatgatatcgaaccagtaagacttgaaga 2866
Db 2886 CCTGCTACCGCTTAAGAAACCGGGGACTAATGATTATAGGCTGTCTCAGGATCTGAGAGA 2945
QY 2867 ggtcaataaagggtgcagaacatacccaagcgtccggaaccctataactcttgag 2926
Db 2946 AGTCAACAAAGCGGTGAGAGCATCCACCCAGCTGCCAACACCTTAAACAACTTGTAG 3005
QY 2927 cgcctcccgctgaaggaacttgtaacaagtatggactttaaagaatccttctctg 2986
Db 3006 CGGGCTCCACCGTCCACACGATGTAACGTGTCTTGAATTAATAATCTTTTTCG 3065
QY 2987 cctagattacaaccactagccaacacttctgcttgccttgcgaatggagagactcaggtac 3046
Db 3066 CCGTGAATCCACCCACACATGATGATCTTCTGCTTTGAGTGAAGAGATCCAGAGAT 3125
QY 3047 gggaagaacgggagctacactggaacccagctgcccgaagggttaagaactcccgac 3106
Db 3126 GGGAAATCTGACGACAAATTAACCTTGACACAGACTCCCGAGGGTTTCAAAAGTCCAC 3185
QY 3107 catcttgacgaagccctcacaggaagccttgccaacttcagatccaacaccctcaggt 3166
Db 3186 CCGTTTGTGAAGCCCTGACAGGAGCCTCGAGACTTCCGATCCAGACCCAGACT 3245
QY 3167 gaccctcccaatgactgatacctgcttcctggggagagcaaccaagaagctgt 3226
Db 3246 GATTTCGCTCACTATGTAGATGACTTACTGCTGGCCGACACTTCTGACACTTGACTGA 3305
QY 3227 agaagtlaggaagcactactcgtgaattgtctgaactaggctacagaacctgtacta 3286
Db 3306 ACAAGTAGCGGGCCCTGTTACAACCTTAGGGACCTCGGATATCGGGCTCGGCCAA 3365
QY 3287 gaaagcccaagatttcagagaagagtaacatacttgggtgatacgttttcggggcgga 3346
Db 3366 GAAAGGCCAAATTTGCCAGAAAGATGATGCTGGGGTATTCCTTAAAGAGGTGA 3425
QY 3347 gcatgtgctgaagcagcaaggaanaactgtagtcagataccgcccacaacagc 3406
Db 3426 GATATGCTGACTGAGAGCCAGAAAGAGACTGTGATGGGGCAGACCTACTCCGAAAGCC 3485
QY 3407 caaacaagtlagagaagttttgggaagactgtgatttgcagacttggatccggggt 3466
Db 3486 TCGACAATTAAGGAGAGTTCCTAGGAGAGGAGGCTTCTGCTGATCCCTGGGTT 3545
QY 3467 tggacttagagagcccccactacccgcttaacaaagaanaagggttctcctgggc 3526
Db 3546 TGCAGAAATGCGAGCCCTTGTACCTCTACCAAAACGGGAGCTCTGTTGAAGGGG 3605
QY 3527 tccctgagcagcaagagcatcttgatgctatcaaaagccctcgtgaagcagcagct 3586
Db 3606 CCCAGACAGCAAAAGGCTTACCAAGATCAAGACGCTCTCTTAACGCTCCGCTG 3665
QY 3587 ggcctccctgacgtaactaaaccttaaccttattgtgatgagcgttaaggaagtagc 3646
Db 3666 GGGATTGACAGACTTACATGAAGCCCTTCGAACTTTTGTGAGAGAAACAGGAGCTAC 3725
QY 3647 cgaagagttttaaaccacaaccttagaccatggagagagcagctgtgctactcgtcaaa 3706
Db 3726 CAAAGGTGCTTACCCCAAAACTGGGGCTTGGGCTTGGGCTGGGCTGCTACTGTCGA 3785
QY 3707 gaagcttgatccctgtagcagtggtgtgcccgtatgtcgaagcctatgcagcgtgtgc 3766
Db 3786 AAGGTAGAACCGTGTGAGCTGGGTGGCCCTTCTCAGGATGTGAGAGCATGCG 3845
QY 3767 catactgtcaagagcgtgacaacttgacttgggaagaatataactgtataagcccc 3826
Db 3846 CGTTCTGACCAAAAGAGCGTGGCAAGCTCACATGGGACAGCATGATCATTTCTGGCCCC 3905
QY 3827 ccatgcatggagaacatcgttcgagagcccccagaacgcatgatagaacaaagccagat 3886
Db 3906 CCATGCAATGAGAGGCACTGATTAAAGCAACCCCTGTGCTGCTCTTCCAGACCCGAAT 3965
QY 3887 gaccacatacaagcgtccttct---cacagaagagtgacttctgcaccacagcgc 3943
Db 3966 GACCCACTACACAGGCTGTGTTCTGACACGGACGACGATTCGAGCAATATGTGCG 4025
QY 3944 tctcaacctgccaactctctgcctgaagaagacttgataccaagtgactcattgtgca 4003
Db 4026 CCTTAAACCCAGGATACCTCTCTCCCTTA---CCTGAGAGGGGGCTGCCAAGATGACCT 4082
QY 4004 tcaactatgattgagagagactgggttcgcaagagccttacaagatacagctgactg 4063
Db 4083 TGACATCTTGCTGGAAGCCACGCACTGACCAAGATCTTAAAGGACAGCTCTCCACA 4142
QY 4064 agaatgtcctaaccgttctactgaagcagactatgtgtgtaaggttaaggaagtgtgc 4123
Db 4143 CGGTGACCAACACTGTGTACAGAGATGGAGCAGCTTCTGCAAAAGGGGACGCAAGGC 4202
QY 4124 tgggcgcaagtggtgagcgggaacccagcatctgggccaagcagctgcgggaaggaac 4183
Db 4203 CGGAGCAGAGTAAACACACGAGACCGAGGTAGTCTGGGCAAAAGCATGCGACCGGGAC 4262

Db	6423	TGGTGCATTCCTTGGCTATCCGGTTTACAAAGCGCGGAAACAGGTCTACTCATGGACA	6482
QY	6286	a-----tgccttaagatagaacgggacagaacccccgtgtgcaatg	6330
Db	6483	ACTGACACATTTGGGGCTTACCTCTTTATGCTCTGACAGACAGCCAGGGCTTACTTCC	6542
QY	6331	ggaccgataaagtaactgctgtacagagggcccccgcccttgagccacgcataactg	6390
Db	6543	GGGATCCGACCTCAAGTTATCAAATCTTAGAGCCCTCGGATCCCAATAGGACCAACCCCGTC	6602
QY	6391	ccgattgcccaatlaaacctgcgtgcgcctgcgatacaacagccgcctlagacaacagtaacc	6450
Db	6603	CTGGCAGACCAACTTGTTGGTTCGGCGTACTAATCCCTTATCCCAACCTGCGCAAGTCTCC	6662
QY	6451	actgattgatctccataccacacgcctagaaactccccagtgltcctgtlaag-----	6504
Db	6663	CCCGCCTCTATGTTGACTCCCACTTGAATTTCCCGGTCCCCACATCCCACTCAGGCCCGC	6722
QY	6505	-----aaagaaagagaaactcagttccatccacagagagctttccaagccataac	6555
Db	6723	CCAGCAGGAACGGGAGACAACTTACTAAATCTTAGTACAGGGGCTTACCAGGCACTCAC	6782
QY	6556	tccaacgacctgtatgcacctctctctgtgtgtgtgtcctactccaaagggcctccat	6615
Db	6783	CTTACCAACCTGATTAATAACTCAAGAGTGTGGTTATGCTTATGTTGTGTGGACCCCCAT	6842
QY	6616	tatgagggatggtctaaagaagaataatcaatlgtaaccaagagcatagaaatcaatgt	6675
Db	6843	TACGAGGGGGTGGCGGCTAGTACTATTATTCACACCACTACTCGTCCGCCAGCTATGCG	6902
QY	6676	acatggyggtcccgaaataaagcttaccctcactgaaagttccggaagggagacatgca	6735
Db	6903	TCCCTGGCCCTCCCAACACAGCTGACCCGTCCGAAGTACAGGAGGGGAGCTCGCATC	6962
QY	6736	ggaagaactcccccatacccaacacacttgctctagaaagtggtttaagcagggc	6795
Db	6963	GGAACTATCCAAAACCTCAACAGGCCCTGTGCACACTACCT-----TAAAGCAAGC	7016
QY	6796	tcaagaataagatattagtaacctgtttaaacaggtgtgtgtgtgtgtatgaatctggttta	6855
Db	7017	AAAGGCTTACTATCTAGTGTGCCCCCACAGGAACTATGTGGGCACTGTACACTGTGATGC	7076
QY	6856	accacctgtttccacactcagctccttaaaccaatcccaagaattctgtgtcatggtccaa	6915
Db	7077	ACTCATGCTTACTGTGCAACCGTCTTATTCACACACTGATGATTTGGTTCTCGTGAA	7136
QY	6916	atgtcccccagatgtactacaacctgtgggaagtggtccttgatgaatagactatgac	6975
Db	7137	TTTATGGCCCAAGGTCACCTTACATCTCTCCACTTACGCTTATAGGCCAATTTGAAAATCC	7196
QY	6976	tataacgaccaaagaagacccgcatccctcaactcagctgtatgtcgtgaattag	7035
Db	7197	CATA---GACATTAAGAAGACACAGTGTCTTAACTTGCGCTTATTTATGATGGCGCTA	7253
QY	7036	acggcgctgtgcgtlaagaacagggacagctgcctccatcacaggaaccaagacgtlaag	7095
Db	7254	ACTATGGGTGGCATCGCGCGGGAGTAAAGGACAGGAACATACCG---CCCTGATGGCCACC	7310
QY	7096	aaaggaactgtgtagcttaactgcgcgcattgaacaagaatctccgaagcttaagagatct	7155
Db	7311	CAGCAGTTTACAGAGCTCCATCGCTGCCGTACAAAGATGATCTCAAGAAAGTGGAAAGATCA	7370
QY	7156	gttaacgacataaagatcccggaactcttctgtcgaagtgtgtctagaacacggagg	7215
Db	7371	ATTACTACTAAGAAAGTCTCTTACTTGTTGCTGAGAGTTGTACTCCAAATTCAGACA	7430
QY	7216	ggaatagatctgtcttctaaagagaagtgaggtlatagtcaccttlaaagaagaatgt	7275
Db	7431	GGCCTAGACCTGTTGTCTTAAAGAGGAGAGCTGTGTGCGCCCTTAAAGAAAGATGT	7490
QY	7276	tgtctctatgtatacaactcaaggagccatcagagactccatgaacaagcttagaaaaag	7355

Db	7491	TGTTTCATGCTGACCATACAGGCTTAGTAGAGTAGATGATGATGGCCAAATTAAGAGAGGA	7550
Oy	7336	ttagaagagcgctgaagaagaagaagcgctgcacagggatggtttgaagaatggttcaac	7335
Db	7551	CTCTCTCAGAGACAAAACATTTTGTATGATCGAGCCAAAGATGGTTGGAAGAGATGGTTTAAAC	7610
Oy	7396	aggtctcccttgatgataccaccctgccttctgcctctagcaggggcccctagtcctgctc	7455
Db	7611	AGATCCCCCGTGTATTACAGGTGATATTCACATCATGSGGSCCTCATATATCTCTTA	7670
Oy	7456	ctgttaacttaacglttgaggccttgcttaatlaalaglttltgtgccttcttagaagaaga	7515
Db	7671	CTAATCTGCTTTTGGACCCCTCATCTTAATCGATTGATTCATTTGTTTAAAGACAGG	7730
Oy	7516	gtgagtgacagtcacagatcagctgacttaggaagaacatgaca	7556
Db	7731	ATCTCAGTAGTCACAGGCTTATGCTCTGACATCAACAATACCA	7771

```

1      RESULT      8
2      US-08-886-642-8
3      ; Sequence 8, Application US/08886642
4      ; Patent No. 5952474
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Pfizer, Abraham
7      ; APPLICANT: Kayman, Samuel
8      ; TITLE OF INVENTION: FUSION GLYCOPROTEINS
9      ; NUMBER OF SEQUENCES: 25
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Fish & Richardson P.C.
12     ; STREET: 45 Rockefeller Plaza, Suite 2800
13     ; CITY: New York
14     ; STATE: NY
15     ; COUNTRY: USA
16     ; ZIP: 10111
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: Diskette
19     ; COMPUTER: IBM Compatible
20     ; OPERATING SYSTEM: Windows95
21     ; SOFTWARE: FastSeq for Windows Version 2.0
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/08/886,642
24     ; FILING DATE: 01-JUL-1997
25     ; PRIOR APPLICATION DATA:
26     ; APPLICATION NUMBER: 08/110,300
27     ; FILING DATE: 20-AUG-1993
28     ; APPLICATION NUMBER: 07/938,100
29     ; FILING DATE: 28-AUG-1992
30     ; ATTORNEY/AGENT INFORMATION:
31     ; NAME: Hone, William J.
32     ; REGISTRATION NUMBER: 26,739
33     ; REFERENCE/DOCKET NUMBER: 0763/010002
34     ; TELECOMMUNICATION INFORMATION:
35     ; TELEPHONE: 212/765-5070
36     ; TELEFAX: 212/258-2291
37     ; INFORMATION FOR SEQ ID NO: 8:
38     ; SEQUENCE CHARACTERISTICS:
39     ; LENGTH: 8323 base pairs
40     ; TYPE: nucleic acid
41     ; STRANDEDNESS: single
42     ; TOPOLOGY: linear
43     ; MOLECULE TYPE: Genomic DNA
44     ; US-08-886-642-8

```

	Query Match	Similarity	21.3%	Score 1747	DB 2	Length 8323
Db	Matches	Local Similarity	55.6%	Pred. No. 0 <td>3005</td> <td>Gaps 27</td>	3005	Gaps 27
	Batched 4586	Conservative	0	Mismatches	3005	Indels 238
Oy	467	tcgcgcgcgcgtcgcacactcttctgcccgtctgtgac--gacgcgcgcgcgtctgcgtgtgt	524			
	498	tcgcgcgcgcgcgtctgcacactcttctgcccgtctgtgac--gacgcgcgcgcgtctgcgtgtgt	557			

Db 2706 GGGCATGGGGCTGGCCGCTTCTGATCACTTGAAGGCAACCTTAC 2765
Qy 2687 accagatcaagtcagacagacagccttgagagagctcgagaagaaatttgccgca 2746
Db 2766 CCCGCTGTCATTAACCAATTAACCCATGTCACAGAGGCAAGTGGGATCAACCCCA 2825
Qy 2747 tcttaagaattatcccaagaggaactcctagttcctgtccaatcccttgaaatactcc 2806
Db 2826 CATACAGAGACTCTGGATCAGGGAATTTCTGGTACCTCCAGTCCCTCGAAGACGCC 2885
Qy 2807 cctgtacacggttagaagccttggaacaaatgatttcgaaccaagtaagagagtaagga 2866
Db 2886 CCTGCTACCCGTTAAGAAACCGGGGACTATGATTAAGCCCTGTCACAGATCTAGAGA 2945
Qy 2867 ggtcaataaaggtgtcagagacatacccaacaggtccgaacaccttaaacctcttag 2926
Db 2946 AGTCAGACAGGGGTGGAGACATCCACCCACCGTGCCCAACCTTACACCTCTTGAG 3005
Qy 2927 cgcctcccgcttgaaacggaactggtacagataggtacattaaagatgctctctg 2986
Db 3006 CGGGCTCCACCGTCCACAGTGTACACTGTGCTTGAATTAAGATGCTTTTTCG 3065
Qy 2987 cctgaattacaccccaactagccaaacttttgcttgaatggaagagatccaggtac 3046
Db 3066 CCTGAGACTCCACCCACACACTGCTCTTGCTTGAAGTGGAGATCCAGAT 3125
Qy 3047 gggaaagacgggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3106
Db 3126 GGGATCTCAGACATTAATTAACCTGACAGACTCCGCAAGGTTTAAAGATGCTTTTTCG 3185
Qy 3107 catctttagcgaagcctcaccacaggaagcagcagcagcagcagcagcagcagcagc 3166
Db 3186 CCGTTGTGATGAAGCCCTGCACAGGAGCTTCGAGACTTCGGATTCAGACAGACGAGCCT 3245
Qy 3167 gacccctccagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 3226
Db 3246 GATTCGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3305
Qy 3227 aagaagtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3286
Db 3306 ACAAGGTACGGGGCCCTTACAAACCTAGGGGACCTCGGATATGGGCTTGCGGCA 3365
Qy 3287 gaagcccaagatttcagagagagagtaacacttggtggaagattgagcgagagca 3346
Db 3366 GAAGGCCCAATTTGGCAAAACAGTCAAGTATGAGGATATCTTAAAGAGGGTCA 3425
Qy 3347 ggaatgctgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3406
Db 3426 GAGATGGCTGACTGAGGCGCAAGAAAGAGACTGTGATGGGCGAGCTTCTCGAAGACCC 3485
Qy 3407 caaacaagtgaagagatttggaagagcagtcagatttgcaagcagtcagtcagtcagtc 3466
Db 3486 TCGCAACATTAAGGAGTTCCTAGGGAGCGGAGGCTTCTGCTCTGTGATGCTTGCGGTT 3545
Qy 3467 tggagccttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3526
Db 3546 TCGCAAAATGAGCCCTTGTACCTCTCAACCAAAAGGGGACTGTGATGAGTGGGG 3605
Qy 3527 tcttgagacacagaaggtcatttgatcacaagaagccctgctgagcgacccctgct 3586
Db 3606 CCCAGACACCAAAAGGCTTACAGAGATCAAGGAGGCTTCTTAACTGAGGCTTGCCCT 3665
Qy 3587 ggcctccctgagcagtaacacacttaacctttagtggaatgagcgtlaagggatagc 3646
Db 3666 GGGATTCGACACTTGAATAGCCCTTCAAACTTTTGTGAGAGAAACAGGGCTAGCC 3725
Qy 3647 ccgagagattttaaaccnaacccttagagacatggaagagacgtgtgtcactactcaaa 3706
Db 3726 CAAGGTGTCTTAACGCAAAACATGAGGCTTGCGCTGCGGCGGCTTACTGTCGCA 3785
Qy 3707 gaagcttgatccctgagcagtggttgccgatagtctgaagagatcagcagtcagc 3766
Db 3786 AAGCTTAAGCCAGTGGAGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3845

Qy 3767 catactgttcaagagacgtgacaaattgactttggagacgaatataactgtatagcccc 3826
Db 3846 CGTTCTGACCAAAAGCGCTGGAGAGCTCACATGGGACAGCCACTGATCATTTCTGGCCCC 3905
Qy 3827 ccatgtatggaagacatggttcgagcagcccccagaacagatgatatgaacagcccgat 3886
Db 3906 CCATGCAATGAGAGCACTGTTAAGCAACCCCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 3965
Qy 3887 gaaccactatcaagcctgctctc-----cacaagaggggtcactttgcctccacacagccgc 3943
Db 3966 GACCCACTTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4025
Qy 3944 tcttaaccctgacacctcttgcctgaagagatgaagacagtgatcatgattgcca 4003
Db 4026 CTTAAACCAAGCTAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4082
Qy 4004 tcaactatgtatgagagacgaggttcgcaagagccttaccagatatccgtgactg 4063
Db 4083 TGACATCTTGGCTGAAGCCCAACGACCTTACAGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 4142
Qy 4064 agaagtgctaaacctggttcagctgaagcagctatgtgtggaagatgaagatgagc 4123
Db 4143 CGCTGACACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4202
Qy 4124 tgggagcagtggtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4183
Db 4203 CGAGACACACTTAACACACGAGACCGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4262
Qy 4184 ttcagcgaacaaagcctgagcctcagcagcagcagcagcagcagcagcagcagcagcagc 4243
Db 4263 ATGCGCCCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4322
Qy 4244 atccataacattataagacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4303
Db 4323 GAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4382
Qy 4304 catctataaacaagaggggttgcttactcagcagcagcagcagcagcagcagcagcagcagcagc 4363
Db 4383 AATATATGAAAGCGCGGGTGTCTCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 4442
Qy 4364 aattcaagcctatagagacacttacttacttacttacttacttacttacttacttacttactt 4423
Db 4443 GATCTTGGCCCTACTGAGAGCTCTCTCTGCGCAAAAGACTTGAATATTCATTTGCC 4502
Qy 4424 tggacatcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4483
Db 4503 GGGACATCAAGAGCGAAACCGCGAGGCAAGGGGCAACAGATGCGGACCAAGCGGC 4562
Qy 4484 caagc 4543
Db 4563 CCGAGAGTACGCCCTAGAGAAATCTCAGAGACTTCCACACTTCTGATTAAGAAATTCAGC 4622
Qy 4544 accc-agaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4602
Db 4623 CCCCTATATCAATGAAATTTTCACTATACGCTGACTGCAATTAAGATCTGACTTAAC 4682
Qy 4603 ctgagactccgagag-----gacctgctataactatctatggaagaaatcctgct 4654
Db 4683 AGGGGCCACTTATGACAGATGCAAGAGAGTGTGGTTTATCAGGGAAGAACCTGTATATGCC 4742
Qy 4655 ccacaagaagaggttagaataatgtccaagagatatacgtctaaaccactaag----- 4708
Db 4743 TGATCAATTCACCTTTAATATTTAGATTTTCTTCAATGATGACCACTCAAGTTTCTC 4802
Qy 4709 aactaaacactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4768
Db 4803 AAAAACAAGGCTCTTTAGAAAGAACTACTCTCTTATATCATCTCAACCGGGATG 4862
Qy 4769 agtggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4828
Db 4863 AAGCTCAAGACATCACTGAGACTTCAAGGCTTGCGACAGTCAATGCGACGCAAGTC 4922

QY 4829 cagaataccctcaggaagaactaaggagagccaccagcgctcactgggaatgga 4888
Db 4923 TGCCTCAACAGAGGAGTTCAGAGGCGCACCGGACCCACCTGGGAATTGA 4882
QY 4889 ctccactgaagtaaaagccggttaataacgaagaataatctatgttttttgaac 4948
Db 4983 TTTCACTAAGGTAACCTGCGCTGTATGGGTATTAATCTTTTGTTCATAGAC 5042
QY 4949 ctttcaagatggtagagcttactactaagaagaagactcaacgttggtgctaa 5008
Db 5043 TTTCTGTGATGGGTAGAGGTTCCACACAGAAAGAACTGCCAAAGTTTGAACAA 5102
QY 5009 gaaatactggagaataatlttccaagaatttgaaactaagtaagtgagacaa 5068
Db 5103 GAACCTACTAGAAAGAAATCTTCCACAGATTGCGCATGCCACAGAGTTTGGGAACGACAA 5162
QY 5069 tggctcagcttcgttgcacaggttaagtcagagagcggcgaagatattggagatgtg 5128
Db 5163 TGGGCTGCGCTCGCTCCAGAGTAACTGACAGAGTACGCCGATTTTACTGGGGTGTGATTG 5222
QY 5129 gaaactgcatgtgcatatacagaccacaaagctcagacagtagagagatgaatgaac 5188
Db 5223 GAACCTACTAGTGGTGTACAGACCCAGAGTTGAGTACAGTGAAGAAAGTGAATGAGAC 5282
QY 5189 catlaaagagacccttaaccaaatgtgacacagagactgcatatgatgtgagctct 5248
Db 5283 AATCAAGAGAGCTTTAACTAAATGTGACGCTTGCACACTGGCTCTAGGGACTGGTCTCTCT 5342
QY 5249 cctgaccttgtgctttttagaggtgagagacaccccttgacagcttggtgagccctta 5308
Db 5343 GCTTCCCTTAGCCCTTATCGAGCCGCGACAGCCGCGGCCCCCATGGTCTCACCCCTA 5402
QY 5309 tgaatgtcctacggggagagcccccgcgttggcagaatctgcttgcataatgtctga 5368
Db 5403 TGAATCTTATATGGGGGACCCCGCCCTTGTAACTTCCCTGATCTGACATGAGGAAA 5462
QY 5369 tggctgcttccacagccttgttctctagagctcaagcgctcgagtgagggagcg 5428
Db 5463 GGTACTCATTAACCCCTCTCTCCAGAGCCCATTTACAGGACATCTACCTGGTCCAGCAGA 5522
QY 5429 agcgtggaagcagctccgggagcgctactcagga-----ggaagactgcaagttccaca 5482
Db 5523 AATCTGAGAGACCGTTGGGCGAGCTTACCAAGAAACACTGGAACCGCGGAGTGGCTCA 5582
QY 5483 tgccttcaaglttgagaatltcagltatgttagagcgccacggctgacgaagaactcgagac 5542
Db 5583 CCTTTCCGAGTGGGTGACACAGTGGGTCCGACAGACCAAACTAAATATCTAAGAAC 5642
QY 5543 tgggtggaagagaccttactctgactttgacacacaaagcgctgtgaagtcgaag 5602
Db 5643 CCGCTGGAAGAGACTTATACCGTCTACGACTACCCACCGCTCTCAAACTGAGCG 5702
QY 5603 aatcccccctgatacagctccaggttaagcggg-----gcacac 5647
Db 5703 CATTGACACGTGATTCACAGCTGCGCCACGTAAGAGCTGCGACACAGATTGAGCGAC 5762
QY 5648 tcccgactcgggttgaagagcgaagaagactgagaatccaccttaagcttcgacctcag 5707
Db 5763 ATCGGAATTCGACATGCGGTTCACGCTCAAAATCCCTTAATAATATGATGACCG 5822
QY 5708 cctgttccttactctaaataactcccaagcgcaagtagtaacgcttataagacgt 5767
Db 5823 CGGACCTCTAATCCCTTAATTCCTCTCTCTCTCAAAAGGGGCCAGATCCGACGA 5882
QY 5768 cgaaccccca-----tagacctatccctactcgtctatattatgacctgatacggt 5823
Db 5883 CCGGCTCCAGCCCTCACAGGCTCTACACATTTACTGGAAGTGAACCAATGGGATGG 5942
QY 5824 gtaactgtaataagcactcggaggtgtgtcctcctagagagcactgttgccgtgaac----- 5878
Db 5943 GAACAGCTATGGCAATATACAGCAACCTCTGTGTGACTTGTGGCCAGTCTCTAC 6002
QY 5879 -----tgacttctgcctcagattgattaaacccgcgtttaaagaacacactccca 5929

Db 6003 CCAGATTTGTATGTTAGTCTCAGTGGGCGCCGACCTGGGGCTAGAGTATCAGGCC 6062
QY 5930 accatgctcggtagtattaggttc-----tatgtgcacagcaagaagaataatc 5985
Db 6063 CCTATTTCTTGCCCCCGGGGCCCCCTTTGTGCTCAGGGAGCGCGGAGCACTTGACGCG 6122
QY 5986 tgggggg-----tttggggaatcctctgtga 6016
Db 6123 TGTCCAGAGACTGACAGAGCCCTTGACCTCCCTACCCCTCGGTGACAACTGCTGG 6182
QY 6017 gatgagctcgtcgtactcccaagatggagactggaatgagcagctctcctcagagac 6076
Db 6183 AACGACTTAACTGACAGCGTACCTCAATTAATCAAGTGAAGGATTTATGTGTGCC 6242
QY 6077 ggttaaatctccttctgttaattccggccggcgaagfacaataatgtgaact----- 6131
Db 6243 GGGTACATGCGCCCGGGAGGCAAGTCTGTGGGGTCCAGACTCTTCTACTGTGCC 6302
QY 6132 -----ataaagaatagagctgtcccatcagacta 6165
Db 6303 TCTTGGGGCTGCGAGACACCGGTAGAGTACTGGAAGCCCTCTCTTTGGGACTAC 6362
QY 6166 gattactaaagaatagcttcaagaaagaaacagaaataatcaaaatgataaa 6225
Db 6363 ATCAGAGTACAGACATCTCCTTACCTTACCGGCTGTTCAGGATTCGAAGACAAATAG 6422
QY 6226 tggtagagcttgggaatagttttatataatggcggggagcagggctcactaac 6285
Db 6423 TGGTCAATCCCTTGCTTCCGTTTTCMAAACCGCGGGAACAGTCACTCATGACA 6482
QY 6286 a-----ttgccttagatagaagaagggagacaccccttggcaatg 6330
Db 6483 ACTGAGACATATTGGGCTTACGCTTATCTCTTGACAGGACCCAGGCTTACTTTC 6542
QY 6331 ggaaccgataaagtaactgtctgaagaagggcccgccggcagcagcagtaactg 6390
Db 6543 GGGATCCGACTAGTTATCAAAATCTAGAGCTCGGATCCCAATAGAACCAACCCGTC 6602
QY 6391 ccgggcccccaataactctgcgtcgagctgacataaacaagccgcttagcaagtaac 6450
Db 6603 CTGGGACACACTTCTCTCCCTCTACCTAATCCCTTACCAACCTGCGACGTCC 6662
QY 6451 actgattgattccttacaacaagccttagaactcccgagtggttctgttaag----- 6504
Db 6663 CCGGCTCTAGTTGAGCTCCACATTTGATTTCCCGGCTCCCACTCCCACTGACCCCG 6722
QY 6505 -----aaggaacagagacttcaagtcacatcccaaggaagcttccaagcactaac 6555
Db 6723 CCAGCAGGAACGGAGAGACAGATTACTAATCTAGTACAGGAGACTTACAGGAGACTCAAC 6782
QY 6556 tccacgacccctgtagcagacttcttctgtgttcttctatcctcagagcctctat 6615
Db 6783 CTTCACCAACCTTGATTAACCTCAAGAGTGGTGTATGCTTATGCTTGGACCCCTAT 6842
QY 6616 tatgaaggaatggtctaaagaagaataatcaatgltaccaaagaagatagaataaagt 6675
Db 6843 TACAGGGGGGTGCGGCTTACGTACTTATTCACACATTAACCTGCCCCACCTAATGTC 6902
QY 6676 acatgggggccccgaataaagcttaccctcaactgaagtttccgggaagggagactgata 6735
Db 6903 TCCGTGGCTCCCAACACAGAGCTGACCTGTCGAAGTACAGGAGGAGGACTGCTGATA 6962
QY 6736 ggaagaactccccaatcccaacaaaccttgcctttagtactggttattagagcagcg 6795
Db 6963 GGAAGAGTCCCAAAACTACACAGGCGCTGTGCAACACTTACCT-----TAAAGCAGGGC 7016
QY 6796 tcgaagaatcagatatttagtactgttataaacaaggttggtgagcagtaactgaggtta 6855
Db 7017 AAAGGGTCTTACTATGTTGCTCCCAACAGGAACATATGTGGGCACTGTAACACTGAGCTC 7076
QY 6856 accccctgtgttccactcagtttcaacaaatcccaagaattctcgtgttcaatgttccaa 6915

Db 1156 GGAGAGTGGCCCTACAGAGAGCCCTTACCCCTTCCCAATGATCCCGCTGCG 1215
Qy 1143 gagcgcagagcagagatcgatatacgcgtgcgcacattagccctccatgcaggg 1202
Db 1216 GGAAGAAAAGAACCCCGGCGGATTTCTACTACTCTCAGGGGTTCCCGCTGCGCTG 1275
Qy 1203 ggcacattgcagccctccatgattgccccttctctcagctctcctaattgnaa 1262
Db 1276 GGAAGGAATGACAGTATCAATCAATCTGCTTCTCTGACCTCTATCACTGAAA 1335
Qy 1263 actaacaccccccttctcagagatccccaacgcctcaacgggttggatgagccct 1322
Db 1336 AATAACAACCCCTCTTCTCGAGAGACCACTAAATTGACGCTTGATGAGATCGGTT 1395
Qy 1323 atgtctcaccagcctacttgatgatgtgtaacaagctgctgcacaacttca 1382
Db 1396 CTCCTACTCATCAAGCCACTTGGATGACAGCAAGCTATTAGGAGCCCTGCTGACG 1455
Qy 1383 accgagagagagagatctgtagagctagaaaaaatgttctctggggccagagg 1442
Db 1456 GGAGAGAAAACAGCGAGTCTCTAGAGCCGAAAGCGGTTCCAGGAGAGAGCGA 1515
Qy 1443 cgaccacgcagctgcagaaatgagatgacatggatctcccttgcagcccggttgg 1502
Db 1516 CGCCCACTCAGCTGCCCAATGACATTATGATGCTTCCCTTGGAACGTCGCGACTGG 1575
Qy 1503 gactacaacgcgtgtagatgagagagcttgaaaactatgcagcagctctggtgcg 1552
Db 1576 GACTACACACCCCAACGAGGTAGAACACCTACTCCATCTCCGCTGCTCTGACG 1635
Qy 1563 ggtctcggggcgctcaagacggcccaatttggcttagatgaaagaggtgtagag 1622
Db 1636 GGTCTCCAAAGCGGGCAGAGCCCAATTTGGCAAGGTAAAGGATTAACCGAG 1695
Qy 1623 ggaacgaacacacctccctcgtaattctctgagagctcagagacacacacacacac 1682
Db 1696 GGACCTATGAGTCTCCTCTCAGCCTTTTACAGAGACTCAAGAGAGCCCTATCGAGATAC 1755
Qy 1683 accccttctgactcctacacacagagcccaagaacacacacacacacacacacacac 1742
Db 1756 ACTCCTTTGACCTGAGAGACCCAGGGCAAGAACCAATGCGCATGTCATTCATCGG 1815
Qy 1743 cagtcgctcgtgatatagaaagaacttcagagacacagagaggtttacagagagcgag 1802
Db 1816 CAGTCGCCCGGATGCGGAGAAAGTTAGAGCGGTTAGAGATTGAGAGATTAAGACC 1875
Qy 1803 ttaacgtactagtagagagagcagagagatglatlaacaagaagagagacagagagag 1862
Db 1876 TTAGAGACTTAGTAGAGAGAGCTGAAGAGATCTTTAATAAGAGAAACCCCGGAGAA 1935
Qy 1863 aaggaacagagaaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1922
Db 1936 AGAGAGAACTATTAGAGAGAGAAACAGAGAGAAAGCAACCCGAGGGCAGAGAT 1995
Qy 1923 gagaagaattgactaagatcttggccgagtggttgaaagagagcagagagagaga 1982
Db 1996 GTGACAGAGAGAGAGAGAGGACCGCAGAAACATAGAGAAATGAGTA-----A 2045
Qy 1983 gagaagattttagaanaattagatcagagccctagacagatcagaga-aacctggcaatag 2041
Db 2046 GTTGTGCTACTGTCTTACCGGGCAGAGACAGAGATGACAGGAGAGAGAGAGAGAG 2105
Qy 2042 gaccacactgcagaaagacagtgctgatttlaaagaanaaagaacacagtggaagaa 2101
Db 2106 GCCCAACTCGACACGACAGTGTGCTACTGCAAAAGAAAGGAGCATTTGGCTAGAGA 2165
Qy 2102 ctgcccacagaagaaacagaagacagaaagtccta-----gctctagaagaaga 2152
Db 2166 TTGCCCCAAGAGAGAGAGACCCCGGAGACACGACCCCAAGGCTCCTCTGACCTT 2225
Qy 2153 taaagattagagagagaggttcgagccctcccgagccagaggttaatttgaaggt 2212
Db 2226 AGACGATTAGGAGGTGAGGGTCAAGAGACCCCGCTGACCCAGAGATTAACCTCTCAGAGT 2285

Qy 2213 ggaagggcaaccagttgattctcgtgtgatacggagagagacattcagctgctaca 2272
Db 2286 CGGGGGGCAACCCGTCACCTCTTACTGTGATCTGGGGCCCAACCTCGTCTCAACCA 2345
Qy 2273 accattaggaanaaataaataaataaataaataaataaataaataaataaataaataa 2332
Db 2346 AATTCCTGAGACCCCTTAAGTACAGTCTGCTGGGTCCAAAGGGGCTACTGAGAGAGG 2405
Qy 2333 gatacgaagactaccgaagaacgcttgacttggatggagcggtaaacacacactglt 2392
Db 2406 GTATCCTGACACCGATGCGCGAGAGTGCATAGCCAGCGATGAGTCAACCTATCTT 2465
Qy 2393 tctgtgcatcccgagtgccagctaccctttagttagaagacttaccagaatggg 2452
Db 2466 CTTCCATGTACAGATTTGCCCTATCTCTGCTAGGAGAGATTTGCTACTTAACATAA 2525
Qy 2453 agctcaaatctcttga--aagaagaacagagagctgtgtgataaanaaaccat 2509
Db 2526 AGCCCAATTCACCTTTGAGGAGATCAGAGACTCAGGTTGGAGCAATGGAGAGCCCT 2585
Qy 2510 cactgtgtgacccctccaattagatgataatgataatatttccccaagtaagcc 2569
Db 2586 GCAAGTGTGACCCCTTAACATAGAAATGATGATCGGTACATGAGACCTCAAAAGCGCC 2645
Qy 2570 tgataagatatacagctc---tggttggagagatttccccaagcctgggcaagaacgc 2626
Db 2646 AGATGTGCTCTAGGGTCCACATGAGCTCTGTGATTTCCCAAGGCTGGGCAAGAACCG 2705
Qy 2627 aggaatgtgttgcaagaacagatccccaagcttactcaactcaagaagcagtgatc 2686
Db 2706 GGGCATGAGGGCTGCGGCTTGCCCAAGCTCTCTGATCATACCTGTAAGCAACCTTAC 2765
Qy 2687 accagatatacgaacagtaaccccttgatgagagagctcagaagaatttgcgcga 2746
Db 2766 CCCGTGTCATTAACCAATTAACCCCATGTGCACAAAGAGCCCAACTGGGATTAAGCCCA 2825
Qy 2747 tgttaagaattataccaacagggcactcagttcctgtgccaactcccttgaatactcc 2806
Db 2826 CATACAGAGACTCTGTGATGAGGAAATCTGTGATCCCTGACGATCCCTGAGAACCGCC 2885
Qy 2807 cctgcacacggttgaagaacgcttgcacaaatgatactcagacagtaacagagcttgaaga 2866
Db 2886 CCGTACACCCGTTAAGAAACCGGGGACTATGATTTAGGCTGTCCAGAGATCTGAGAGA 2945
Qy 2867 gtlcaataaagagtgacagacatatacccaacgltccccaacccctataacactctgag 2926
Db 2946 AGTCAACAACCGGGTGAAGACATCCACCCGACCGCTTACCAACCTCTTGTAG 3005
Qy 2927 cgcctcccgcccgaaacgagacgtgatacagatgagacttaaaagatgaccttctgt 2986
Db 3006 CGGGCTCCCACTGTCACACAGTGTGACTGTGACTTAAAGATGCTTTTCTG 3065
Qy 2987 cctgagatlaaccccaactagcaacacacttcttgcctcgaatgagagatcagagtlac 3046
Db 3066 CCGAGAGACTCACCCACAGCAGTCACTGCTCTTCCGCTTTGAGTGAAGATCCAGAGAT 3125
Qy 3047 ggaagaacccggcagctcaactggaacccagctgccccaaaggttcaagaactccccgac 3106
Db 3126 GGAATCTCAGCAATTAACCTGACACAGACCTCCGAGGGTGTCAAAAGATGCTCCAC 3185
Qy 3107 catcttgaagagccctacacagagagacttggccaacttcagatccaacacccctcaggt 3166
Db 3186 CCGTTTGTAGAGCCCTGACAGGAGCTTCGAGACTTCGAGATCCACACCCAGACT 3245
Qy 3167 gaccctccctcagtagctgtagactgtcttgcggggagagcaacaaacagagactgtc 3226
Db 3246 GATTTCGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3305
Qy 3227 agaaagtlacgaagacactagctgtaattgttgcactaggtacagaagcctctgctaa 3286
Db 3306 ACAAGGTAGCGGGCCCTGTATCAAAACCTAGGGAGACTCGGATATCGGGCTCGGCCAA 3365

Db 5523 AGTCTGGAGACCGTTGGCGGACGTTACCAAGAACAACTGGACGGCGGTAAGTGCCTCA 5582
 QY 5483 tgcctccaaagttaggaattcagctatgttagagccacgctgcaggaacctgagac 5542
 Db 5583 CCCCTTCCGAGTGGGTGACACAGTGGGTCCGACAGACCAACATAAATCTTAAGACC 5642
 QY 5543 tgcgttagaaggagcctatctgttactttgaccacacagcgctgtagaagtcgaag 5602
 Db 5643 CCCGTGGAAGAGACCTTATACCGTCTACTGACTACCCACCGCTCTCAAAATGAGACG 5702
 QY 5603 aatccccaacctgtagcattcattccacggttaagccggc-----gcacac 5647
 Db 5703 CATTCGACGCGGATTCACCGCTGCCAGTAAAGGCTGCGACACAGATGAGCCACG 5762
 QY 5648 tcccgactcggtgtgagagccgaaagactggaaatcccttaagcttcgctccatcg 5707
 Db 5763 ATCGGAATCGACATGCGGTTCACACGCTCTCAAAATCCCTTAAGTAAGATTGACCCG 5822
 QY 5708 ccttgcttctacttaacaataactcccgaggccagtaglaagccttatagacagt 5767
 Db 5823 CGGAGACTCTTAATCCCTTAATTTCTTCTGCTCTCAAGGGGCGACATCCGACGCA 5882
 QY 5768 cgaaccccca-----tagaccttatacccttaccctgctgataatgaacctgatacgg 5823
 Db 5883 CCCGGCTCCAGCCCTCACAGGCTTACACATTAACCTGGGAGTGAACCAATGGGGATCGG 5942
 QY 5824 gtacctgataatagcactcgagtggtgctcttagagagccctgtgagcctgaac----- 5878
 Db 5943 GAGACATATGGGCAATATGAGGCAACACCCCTGTGTGACTGTGGTGGTGGCAGTCTCACC 6002
 QY 5879 -----tgcaatttgcctccgattgataaccgctgtgtaaaagacacctccca 5929
 Db 6003 CCATATTTGTGTATGTATGATGCTCAGTGGGCGCCGACAGGGGGCTAGAGTATCAGGCC 6062
 QY 5930 acctaagtcogtagttagtggtc-----tatgtgcacagacacagagaagaagaaatag 5985
 Db 6063 CCCATTTCCTCGCCCGGGGGCCCTTGTGTGCTCAGGAGGACAGCGGGAACGTTGACGGC 6122
 QY 5986 tgttgagg-----ttctgggaatccctctgttaga 6016
 Db 6123 TGTCCGAGACACTGCACAGCCCTTGACCTCCCTCACCCTCGGTGACACACTGCTGG 6182
 QY 6017 gatgagctgcgtacactccaacgtagagactgaaatgagccgactctccagagac 6076
 Db 6183 AACAGACTTAACTGAGACCAAGTAACTCATTAATCAAGTAGAGGATTTTATGCTGCCCC 6242
 QY 6077 gggtaaaaattctcttgcataatccgcccgggcaagtagtaaaatgataaact----- 6131
 Db 6243 GGGTCACATGCGCCCGGGAAGCCAAAGTCCGTGTGGGTCCAGACTCTCTTCTACTGTGCC 6302
 QY 6132 -----atataagataagagctctccccaacagactta 6165
 Db 6303 TCTTGGGCTCGAGACACAGCGTAAAGATATCTGAAAGCCCTCTCTTCTTGGGACTTAC 6362
 QY 6166 gattactaaagataagttcaactgaaagaaacaggaataatctcaaaagtgtataaa 6225
 Db 6363 ATCAGATAGAGCAACAATCTACCTCTAACCGAGCTGTTCAGTATGCAAAAGACAAATAG 6422
 QY 6226 tggatagagctggggaatagtttttatatataagcgaggagagagagcaacttaac 6285
 Db 6423 TGGTGCAATCCCTTGCTATCCGCTTACAAACGCGCGGGAACAGGTCACTCTATGAGCA 6482
 QY 6286 a-----ttcgcccttagaataagacgaggaagaaacccctgtgcaatg 6330
 Db 6483 ACTGAGACATATGGGCTCTAAGTCTTATGTCTCTGAGACGAGCCAGGCTTACTTTC 6542
 QY 6331 ggaaccgataaagtaactgctgaacaggggccccggccctggagccacagcataactg 6380
 Db 6543 GGGATCCGACTCAGTTATCAAAATCTAGACCTCGGATCCCAATAGGACCAAAACCCGTC 6602
 QY 6391 ccggtgcccacataactgctgcgagctacataacagcgccagtagaacaagtac 6450
 Db 6603 CTGGGAGACCAACTTCTTCCGCTTACTTAATCCCTTACCCAAACCTGCCAAGTCTCC 6662
 QY 6451 actgattgattcctaccacacgacctagaactcctccagggtgtctctgtaag----- 6504
 Db 6663 CCCGCTCTAGTTGACATCCCATTTGAATTTCCCGTCCCGACATCCCATCAGGCCCG 6722
 QY 6505 -----acagagacagactctcagctcatalcagaggaagctttccaaagcataac 6555
 Db 6723 CCAGCAGAGACGGGAGAGATTAATTAATAGTACAGGAGCTTACAGGCACTCAAC 6782
 QY 6556 tccacagacccctgtagcactctctctgttggtctgtctatccctcagggcctccat 6615
 Db 6783 CTTACCAACCCGTAAATCTCAGAGTCTGTGTATCTCTAGTCTGAGACCCCGCTCA 6842
 QY 6616 tatgagggatggtctaaagaaatccaatgtagcaaaagagacatagaatcaatgt 6675
 Db 6843 TACGAGGGGGTGTGGCGTCTAGTACTATTTCACACCATTAACCTGCCCCAGCTAACGC 6902
 QY 6676 acatgggggtcccgaaataagcttaacctcaactgaagttccgggaaggaacatgata 6735
 Db 6903 TCCGTGGCTCCCAACACAGCTGACCTGTCCGAAAGTGTGACGCGGAGCTGCTA 6962
 QY 6736 ggaagagctcccccacccacacacacttgcataagtagttagttagttagagagcc 6795
 Db 6963 GGAACAGTCCCAAAATCTACACAGCCCTGTGCAACACTACCT-----TAAGCAGGC 7016
 QY 6796 tcaagaaatcagatttagttagctgtgataacaggtggtggaatgaataactggtta 6855
 Db 7017 AAGGGTCTTACTATCTAGTACTGCCCCACAGAACTATGTGGGCAATGATGACCTGACCTC 7076
 QY 6856 acccctgtggttccactcaactcaacttaacacacacaaagtttctgtgtaactgta 6915
 Db 7077 ACTCATGCTATCTGCGCACCGTGTATATCCACACCTGACTATGTGGTCTCTGTA 7136
 QY 6916 atcgtcccccagtagtactacatccctgaagagtggtctgtatgaatataactatg 6975
 Db 7137 TTATGGCCAGGGTACACTACCATCTCCCACTTACGTATAGCCAGTTGAAATATCC 7196
 QY 6976 tatbaaccgacaaagagaagacccglatcccttaacctagctgttaaigtctgagtagg 7035
 Db 7197 CATAT--GACATTAAGAAAGAACACAGTGTCTTACCTTGGCTTATTTAAGTGGGCTA 7253
 QY 7036 acggcctgtggttaggaagacagagctgacccgtatcacagagacacagagtagag 7095
 Db 7254 ACTATGGGTGATGCCCGCGGAGTAGGACAGACATACCG-----CCCTGCTGCCAC 7310
 QY 7096 aaagactgtgtagtatactgacacagacatgacagagatccagagccttaagagctc 7155
 Db 7311 CAGCAGTTTCAGAGAGCTCATGCTGCGGTACAAAGATGATCTCAAGAGTCAAGAAAGTCA 7370
 QY 7156 gttagcaacctagaagagctccctgactcttctgtgaaagtgttctcaagacggagg 7215
 Db 7371 ATTACTTACCTTAAAGAGTCTTACTTGTGTGAGGTGTACTGAGATGAGACGA 7430
 QY 7216 ggaattagctgctgttcttaagagaagtggtttagtgcagccttaaaagaaatgt 7275
 Db 7431 GGCCTTACAGCTGTGTCTTCTAAAGAGGAGGAGCTGTGCGCTTAAAGAAAGATGT 7490
 QY 7276 tgcctatgtagatactacacagagcactcagagactccaatgaacagcttagaagaaag 7335
 Db 7491 TGTTCATAGCTACCATACAGGCTTAAAGATAGATAGTATGCAAAATTAAGAGAGA 7550
 QY 7336 ttagaagagcttcgaaggaagagagctgacacaggggtgtgtgaagagaggtcaac 7395
 Db 7551 CTCTCTCAGAGACAAACTATTGTGTGAGCGCAAGATGAGTGTGAAAGATGTTTAA 7610
 QY 7396 aggtcccttgtagtgaacacccctgtcttctgctgaacggggcccttagtagtctgtac 7455
 Db 7611 AGATCCCGCTGTTTACACAGTTGATATCCACATATGAGGGGCTCTCTATATACCTTA 7670
 QY 7456 ctgtacttaacagtggtgacctgtcttaataaagttgttgaccttgtagagaagaa 7515
 Db 7671 CTATTTCTGCTTTTGTGACCTGCTATTTAATCGATGATGTTCAATTTGTAAAGCAGG 7730

Db 5829 AGAGGGAACGTATACGAGAGGAACAGAGGAAAAAGAACGCCGTAGGACAGAGCAT 5770
QY 1923 gagaagaatttgactaagaatcttgcgcagtggttgaaggagaagcagcagggagaga 1982
Db 5769 GAGCAGAAAGAGAAAGAAAGATGATGAGACATAGAGAGATGAGCA-----AGCTA 5716
QY 1963 gagaagaatttgagaataatagtcagccctagcagcagcagcagcagcagcagcagc 2042
Db 5715 TTGGCCACTGTCGTAGTGAAGCAAGACAGATAGACAGGAGAGAGAAC-----GAAGGAGC 5659
QY 2043 acccaactcgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2102
Db 5658 TCCCACTCGATCGGACGACCTGTCCTACTGCAAGAAAGGGGCACTGGGCTPAAGAT 5599
QY 2103 tgcaccaaga-----gggaaacaagaagcagcagcagcagcagcagcagcagcagcagcagc 2153
Db 5598 TGTCCCAAGAAACACGAGGACCTCGGGAGCCAGAACCCAGACCTCCCTCCTACCCCTA 5539
QY 2154 aagaattgggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2213
Db 5538 GATGCTAGGAGGAGTCAAGGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5479
QY 2214 gaggagcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2273
Db 5478 GGGGGGCAACCCGTCACCTTCTCTGATAGATAGTGGGGCCCAACCTCGCTGATCCCAA 5419
QY 2274 ccaataggaaactaaagaaactaaactcctggtgagtgagtgagtgagtgagtgagtgag 2333
Db 5418 AATCCTGGAGCCCTAAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT 5359
QY 2334 tatcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2393
Db 5358 TATGCTGAGACACGATGAGTGAAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGT 5299
QY 2394 ctgttcacccctgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 2453
Db 5298 CTCATGATGACGAGACTGTCCTATCCTCTGATAGAGAGATGTCGATGATGATGATGATGAT 5239
QY 2454 gctcaaatctcttga---acaagaagcagcagcagcagcagcagcagcagcagcagcagc 2510
Db 5238 GCCCAATTCACCTTTGAGGGGATGAGGAGTCAAGGAGTATGAGGAGTATGAGGAGTATGAG 5179
QY 2511 actgttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2570
Db 5178 CAATGTTGACCTTAATATATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT 5119
QY 2571 gataaagataacagtc---tggttgagcagcagcagcagcagcagcagcagcagcagcagc 2627
Db 5118 GATGTTCTCTAGGGTCCACATGCTGCTGATTTTCTCAGGGCTGGGGGAAACCCGGG 5059
QY 2628 gggatgggttggcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2687
Db 5058 GGGATGGGAGTGGGAGTGGGCAAGCTCTGATCATATACCTCTGAAGAGCAACCTCTAC 4999
QY 2688 ccaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2747
Db 4998 CCCGTGTCATTAACATATACCCATCTCACAGAAAGCCAGACTGGGAGTCAAGCCCCAC 4939
QY 2748 gticaaagatataccaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2807
Db 4938 ATACAGAGACTGTTGAGCAAGGGGCTACTGTAACCTGCGAGTCCCTCTGGAACAGGCC 4879
QY 2808 ctgctaccggttagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2867
Db 4878 CTGTATCCCTGTTAAGAACAGGAGTAAATATATAGCTGTGCTGAGATGTGAGAGAA 4819
QY 2868 gtcaataaaggtgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2927
Db 4818 GTCAACAAGGGGAGGAGTCAATCCACCCAGTGGCCCAACCCCTTACCAACCTCTGAGC 4759
QY 2928 gccctccgcctgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2987
Db 4758 GGGCTCCACCGTCCCAACGAGTGTATACATGCTGATTTAAGAGATGCTTTTCTG 4699

QY 2988 ctgagattacaccoccatagccaaccatttgccttcgaaatggagagatccaggtacg 3047
Db 4698 CTGAGACTCCACCCACACACTGACCTCTTGTGCTTTGAGTGGAGATCCAGATG 4639
QY 3048 ggaagaacccggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3107
Db 4638 GGAATCTCAAGACAAATTTGACTGAGACAGCTCCACAGGGTGTAAAAACAGTCCAC 4579
QY 3108 atctttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3167
Db 4578 CTGTTGATGAGGCTGACGACAGAGACTGAGACTTCCGATCCGATCCAGACCCAGCTTG 4519
QY 3168 accctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3227
Db 4518 ATCTGCTACAGTACGATGAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4459
QY 3228 gaaagtaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3287
Db 4458 CAAGGTACTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4399
QY 3288 aagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3347
Db 4398 AAAGCCCAATTTGCGAGAAACAGTCAAGTATCTGGGGTATCTTAAAGAGGGTCA 4339
QY 3348 cgaatgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3407
Db 4338 AGATGGCTGACTGAGGCGCAAGAAAGAGACTGTGATGGCGAGCTTACTCGAAGACCTT 4279
QY 3408 aagaagtaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3467
Db 4278 CGACAACCTAAGGAGTCTCTAGGAGCGAGGCTTCTGCGCTGCTGCTGCTGCTGCTG 4219
QY 3468 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3527
Db 4218 GCGAATATGACGACCCCTTGTATCTCTACACCAAGCGGAGCTGTTATATGAGG 4159
QY 3528 cctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3587
Db 4158 CCAGACCAACAAAGGCTATCAAGAAATCAAGCAAGCTCTTCAATGCCCCACCGCTG 4099
QY 3588 gccctcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3647
Db 4098 GGGTTCACAGATTTGATGAGCTTGTGAACTCTTTGAGAGAGCAAGCAGGCTACCC 4039
QY 3648 cgaagagcttaacccaacccctagacacatggaagagcagcagcagcagcagcagcagcagc 3707
Db 4038 AAAGGCGTCTTAACGCAAAAGCTGGAGCCTTGCGCTGCGCTGCTGCTGCTGCTGCTG 3979
QY 3708 aagcttgatcctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3767
Db 3978 AAGCTGAGACCCAGTGGCACTGGCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3919
QY 3768 atactgtaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3827
Db 3918 GTTCTGACAAAGATGTGGCAAGCTCACTATGGAACGCGTGTGCTATTCCTGCGCCC 3859
QY 3828 catgcatggaagacatcgtcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 3887
Db 3858 CATGCCGTGAGAGCACTAGTAAAGCAACCCCTGATTCCTGCTGCTGCTGCTGCTGCTG 3799
QY 3888 acccaatcaaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3944
Db 3798 ACCCATTAACAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3739
QY 3945 ctcaacccctgacacctctctgctgtaagaagcagcagcagcagcagcagcagcagcagcagc 4004
Db 3738 CTAAATTCAGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3682
QY 4005 caactatgagtgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4064
Db 3681 GACATCTTGCTGAGAGCCCAAGCACTAGATCAAGTATTTACGAGACAGCCCTCCAGAC 3622

QY 4065 gaatgtcctaactgtgttctactgaagcaagcactatgtgtggaagtaagatggtc 4124
DB 3621 GCCGACCCACACCTGGTATACCGATGGAGAGCTTCTGCAAGAGGCGACGTAAGGCC 3562
QY 4125 gggggcgaatgtgtggaagcggcgcacgactctggccgagcagcctgcgggaagact 4184
DB 3561 GGAGGACGGGTGACCACTGAGACTGAGTAATCTGGGCGAGGGCATTCGCGCGGACA 3502
QY 4185 tcaagcacaagcctgaagcctatgagccctcaagcagccttgcggctgcggcaagga 4244
DB 3501 TGGGCCCAAGAGCTGAACTGATAGCGTCAACGCCCTTAAGATGCGCAAGAGTAG 3442
QY 4245 tccataacattatacgaagcagcagatgactcttgcgactgaacgctacagcggcc 4304
DB 3441 AAGCTAAATGTTTATCTGATGATGCGCTTACCTTTGGCACCGCCCATATTCATGAGAA 3382
QY 4305 atctaataacaaagggtgttcttaccctcaagagggaagaaataaagaagaagaa 4364
DB 3381 ATATTCANAGAGCGCGGGTGTCTACATCAGAGAGAAAGATCAGACAGACGAG 3322
QY 4365 attctaagcctattagaagccttactatgtccaaagcctagctatatacctgtcct 4424
DB 3321 ATCTTAGCCCTACTAAAGGCTCTTCTTGGCCAAAGACTTAGATATTCATTTGCCG 3262
QY 4425 ggaactcagaagcccaagatctcatctagagggaacagcagatgctgaccgggtgccc 4484
DB 3261 GGACATCAAAAAGGAAACAGGCGACAGGCGCAACCGGATGGCCGACCAAGCGGCG 3202
QY 4485 aagcgcgagcccgagcgctga-----acccttgcctataatagaagcccaagc 4537
DB 3201 CGAGAGTAGCCACTAGAGAAATCTCAGGAACCTTCACACTTCTGATGAGAAATCTCAAC 3142
QY 4538 cccaagaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4597
DB 3141 CCTATACCCATGAAACACTTTTACATACACTACTGACAAAGAGATTGACCAACTA 3082
QY 4598 g--ttctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4655
DB 3081 GGAGCCACTTATGACAGTGCAGAAATATTTGGTCACTCAAGGAAAGCCGTATATGCT 3022
QY 4656 cacaagaagggttagaataatgtccaaagaataatgtccaaagaataatgtccaaagaata 4714
DB 3021 GATCAATTCACCTTTGAGTACTAGACTTTCTTCAACATTTGACCCACTGAGCTTCA 2962
QY 4715 aacactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4774
DB 2961 AAAAAGAGGCTCTCTCTAGAGAGAGAGCCCGCTCCCTACTACTGCTGAACCGGAGTGA 2902
QY 4775 tgaactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4834
DB 2901 ACACTCAGAAATATCATCTGAGACTGCAAAAGCTTGTGACAAAGTCAATGCGACAGTCT 2842
QY 4835 accctccaggaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4889
DB 2841 GCCCTTAAAGCAAGAACTAGGCTCCGCGGCGATGCGCTGCGACACACTGGGAATGAT 2782
QY 4890 ttcactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4949
DB 2781 TTTACCGAGGTAAACCTCGGATTTGATGCGTATAGATATCTTTAGTTTGTGATGATCT 2722
QY 4950 tttcagaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5009
DB 2721 TTTTCTGGCTGGATGAGAGCTTTCCCAACTTAAGAAAGAAACCGCGCAAGGCTGATGACAA 2662
QY 5010 aaatctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5069
DB 2661 AAACCTGAGAGAGATCTTCTTAAAGTTGCGCATGCGCGAGGATTTGGGAACCTGCAAT 2602
QY 5070 ggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5129
DB 2601 GGGCGTGGCTCTGCTCCCAAGGTGATGACATGCGCGCATCTGTTGGGATTTGATGAG 2542
QY 5130 aaactgcaatgtgcatalacagaccccaagcctcagcagcagcagcagcagcagcagc 5189

DB 2541 AAATTAACATGTCATACAGACCCCAAGGCTCAGCTAGCTAGCAAAAGATATAGAGCC 2482
QY 5190 attaaagagcccttaccatattgacacagcagcagcagcagcagcagcagcagcagc 5249
DB 2481 ATCAAGAGAGCTTTAACTTAATTAACGCTTGAACATGCGCTTGAAGACTGGTGTCTCTTA 2422
QY 5250 ctgccttctgtcttctttagagcagcagcagcagcagcagcagcagcagcagcagcagc 5309
DB 2421 CTCCCTTAGCCCTGATACAGAGCGCGACAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 2362
QY 5310 gaattgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5369
DB 2361 GAGATCTTAATATGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2302
QY 5370 gtgctgtcttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5429
DB 2301 GTTACTACAGAGCCCTTCTCAAGCTCACTTACAGGCGCTTCAAGCGATACAGAGAG 2242
QY 5430 gctgtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5483
DB 2241 GTCTGAGAGCGCAGCTGCTGGCGCTATCAGAGACAGCAACGACGACGATACACAG 2182
QY 5484 cgtctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5543
DB 2181 CCTTCCGCTGCGCGCACACCGTGTGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2122
QY 5544 cgtgtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5603
DB 2121 CGTTGAGAGAGCGCTTATACCGTCTGCTGACACCGCGCGCGCGCGCGCGCGCGCGCG 2062
QY 5604 atcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5663
DB 2061 ATGCTGCGCTGATACAGCGCGCGCTCAAGTAAAGCGCGCGCGCGCGCGCGCGCGCG 2002
QY 5664 agagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5723
DB 2001 GCATCAGAGACGACATGAGAGGCTCC--ACCGTTCCTCAAAACCGCTTAAAGATATGA 1945
QY 5724 aacataactcctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5783
DB 1944 ACCCGTGGGCGCGCGCTGATGCTCTGGGATCTTAATTAAGGAGAGATCATGATCAAC 1885
QY 5784 tt--talccttaccctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5841
DB 1884 ATGACAGCGCTTACCGAGGCTTCAATGTTACTTGGAGAGTTACCAACTTAATGACAGGAC 1825
QY 5842 cgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5901
DB 1824 AAACAGCTAACGCTTACCTCTCTGGGAGCATGACAGATGCTTCTCATGCTGACT 1765
QY 5902 aacccgcgtttaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5961
DB 1764 TCGACTTGTGCGATTTAATAGGAGGAGATTGGGATGAGCTGAGCTTGGGTGCGCAC 1705
QY 5962 ccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6014
DB 1704 CCGGGGAGAGAAACGCGGAGAGACTTTGACTTCTATGTTGGCCCGCGGACTGATGAC 1645
QY 6015 gagaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6071
DB 1644 CAACAGAGGCTGAGAGGCGGAGAGAGGCTACTGTGCAAAAGGCGCTGAGACCACTG 1585
QY 6072 ggaacgcgttaaaattctctcttcttcttcttcttcttcttcttcttcttcttcttctt 6131
DB 1584 GACAGGCTATCTGAGAGCCATCATCATGAGGAGCTTAATTTCCCTTAAAGCAGAGAA 1525
QY 6132 atataagataagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6187
DB 1524 CCCTGAGATACAGGCGCGCTGATGATGCTTCTCACTGCTGCTGCTGCTGCTGCTGCTGCT 1465
QY 6188 ctgaaggaagaaacagaaataatctcaaaagcagcagcagcagcagcagcagcagcagc 6247

Dh 678 GAGGGATGCTCACACACAGTCGGTAGATGTCAAGAGACAGCTTGGTTACCTTCG 737
Qy 705 gctctgaatgccaacatlcgatgttgatgacatcaagggagaccttaattcga 764
Dh 738 TCTGAGAGATGGCCACACTTTAAGTGGATGGCCGAGAGACGACACTTTAACCGAGC 797
Qy 765 attactgctgttaaggcaatcattttcaagactgacccggtctcatalcctgaacag 824
Dh 798 CTCAATCACCCAGTTAAGATCAAGTCTTTTTCACCTGGCCGCGATGACACCCAGACAG 857
Qy 825 gagccatataccttaagtgtgcaagatttgcagaagaatcctccgcatgtggttaaca 884
Dh 858 GTCCCTTACATGTGTACTGGAGCCCTTGTGACCCCTCCCTGGGTCAACCC 917
Qy 885 t-----ggtcaataaacaagaagaccaggttcctccgaatcctgac 925
Dh 918 TTTGTACACCTTAAGCTCCGCCCTTCCTTCTCATACCCGCCGCTTCTCCCTTGAA 977
Qy 926 tcttgagagaaaaacaacacatcgcgcgaaaaaagtcagccctctcgtatcacc 985
Dh 978 CTTCTGTTGACCCCGCCCTGATCTCCCTTATCAGCCCTCACTCCTTCTGTAGGC 1037
Qy 986 ccgagatcagagagacgcgcgaacttgcgcgaaccccaacctgttcccaacccctatc 1045
Dh 1038 GCCAACCCTTAACCTCAAGTTCTTTCTGACAGTGGGGGCGGCTCATGACCTACTTACA 1097
Qy 1046 cagacagaggtgtgtgtagagagacctctgccccctcgtgagctccggtgtgtagagacc 1105
Dh 1098 GAATACCCCGCCCTTATAGGAGCCCAAGACACCCCTTCCGACAGGAGGAAATGTT 1157
Qy 1106 tgcgtccggaactcgcagcgcgagagcgccaccgcgagcg----- 1147
Dh 1158 GGAAGAACGACCCCTGCGGAGAGGACCGAGACCCCTCCCAATMGATCTGCCATAGT 1217
Qy 1148 gaacagagagatcgcatatctacgcgcgc-----gaacctatgccccctccatgcagag 1202
Dh 1218 GGGAGAGGGAGCCCTGTGGGAGCTCACTACCTCGAGGACATTCCTCCCGCGCA 1277
Qy 1203 ggcgaatctgacgccccctcagatlttgcccccttctctcgtcagatcctataattgaa 1262
Dh 1278 GGAAGAAAGGACAGCTTCAATAGTGGCGCTTCTCCTTCTGACCTTTTCAACTGGA 1337
Qy 1263 actacaatcccccttctcgcagagatcccccaacgcctcaaggggtgtgtagagctcc 1322
Dh 1338 AATATATACCTTTTCTTGTGAAGATCCAGTAACTGACACTGTGATCGATCTGTT 1397
Qy 1323 atgtctcagcagcacttctgagatgtgtcaacagctgtgtcagacatcctca 1382
Dh 1398 CTATTCACCAATAGCCCACTGGAGACTGTACAGCTGTGGGAGCTCTGTGACC 1457
Qy 1383 accgagagcagagagaaatctgttagagctagaaaaaatgttccctgagcgcagag 1442
Dh 1458 GGAAGAAAGAAACAGCGGTGCTCTTAGAGCTAGAAAGCGGTGGGGCGCATGTATGG 1517
Qy 1443 cgaacacgcagttgcaaaatgagatgacatggaatttcccttgaactgcgcgcgtttg 1502
Dh 1518 CGGCCCACTCACTGACCAATGAAGTGAAGCGCGCTTTTCCCTCAGCGCCCAAGACTG 1577
Qy 1503 gactacaacacagctgaagtagagaggttgaaaaaatcatgccaagctcgtgcgcgcg 1562
Dh 1578 GATTACACACCCAGGAGGTAGGAACCACTATGTCCACTATGCGCAGTTGCTCTAGGG 1637
Qy 1563 ggtctccgagcctcaacagcgcacactaattgtgctaagtaagagaggtgtatgcag 1622
Dh 1638 GGTCTCCAAAGCGGGGAGAAAGCCCAACCAATTTGGCCAAAGTAAGAAATTAACAA 1697
Qy 1623 ggcagcagaacactccctcgtgatcttctgtagagctcatggaagccttcaagcggttc 1682
Dh 1698 GGGGCGCAATGATGTCTCCCTGCGCTTCTTAGAGAGACTTAAGAAACCTATGCGCAGTAC 1757
Qy 1683 acccctttagctactacactcagagagcccaagaagcctcaagccttgccttcatgttg 1742
Dh 1758 ACTGCTTATGACCTGAGGAGCCAGGCGCAAGAACTAATGTGTATGTCTTCAATTTGG 1817

Qy 1743 cagtcgctcgtgatctcaggaagaacttcagagactgtgaaggttctacagagcttg 1802
Dh 1818 CAGTCTGCCCAACATTTGGAGAGATTAGAGAGGTTAGAAATTTAAAAACAAGAGC 1877
Qy 1803 ttaacgtatcattgagagagagagaggttctatcaagaaggagagagagagagag 1862
Dh 1878 CTTGAGATTTTGGTTAGAGAGGCGAAGAAATGATTTTAAATAAAGAAACCCCGAGAA 1937
Qy 1863 aaggaacagagaaagaaagagagagagagagagagagagagagagagagagagag 1922
Dh 1938 AGAGAGAGAGATGTATCAGAGAGAGAAACAGAGAGAGAGAGAGAGAGAGAT 1997
Qy 1923 gagaagaatttgtaagatccttgcgcgaagtggtgaagggagagagagagagagag 1982
Dh 1998 GACACAGAAAGAGAGAGAGAGATCTGTAGAGACATATGAGAGATGAGCA-----AGCTA 2051
Qy 1983 gagaagattttagaanaattagtcagagcccttagacagctcagagagacctggcataag 2042
Dh 2052 TTGGCCACTGTCTGTAGTGGACAGAAACAGATAGACAGAGAGAGAGAGAGAGAG 2108
Qy 2043 acccctcgcagacagagacaggtgtgcgtatgttaagaaagagacactggcagagagac 2102
Dh 2109 TCCCAACTGATCGCGAGACAGAGTGTGCTACTCAAAAGAAAGAGGCGACTGGCTTAAAGAT 2168
Qy 2103 tgcgcgaagaa-----gggaacaaagagcgcgaagtccttagctctagaagaaga-----t 2153
Dh 2169 TGTCCCAAGAAACACACAGAGAGCTTGGGGACCAAGACCCCAACCTCTCCCTGACCTTA 2228
Qy 2154 aagaattagagagagagaggttgcgacccccctccgcagcagagtaacttgaaggtg 2213
Dh 2229 GATGATAGGAGAGTGTAGAGGTGAGAGAGCCCGCCCTGAACCCAGAGATACCTCAAAATG 2288
Qy 2214 gaggggcaacaggttgaagttcgtgtgtatcagcagagcggagacatcagtgcttaaca 2273
Dh 2289 GGGGGGCAACCCGTAACCTTCTGTGATGATGGGGGCCCAACACTCCGCTGACCCAA 2348
Qy 2274 ccattagaaactaaagaaagaaatccctgtgtgtatgtgtgtgcacagggagaaagag 2333
Dh 2349 AATCCGAGACCCCTATAGATATAGTCTGCTGGTTCACAAAGGCGTACTGAGAGAGAGG 2408
Qy 2334 tatcatagactaccgaagacaggttgaacttgcgtgtgtgtgtgtgtgtgtgtgtgtgt 2393
Dh 2409 TATCGCTGACCAAGATGAGCAAGATGATAGTACCGGATGAGTCAACCATCTTTC 2468
Qy 2394 ctgtcatccctgagtgccagttaccccttcaagttagagacttacttcaagaatgagga 2453
Dh 2469 CTCATGTACAGACTGTCCCTATCTCTGTTAGAGAAAGATTTCTGACTTAACATAAA 2528
Qy 2454 gctcaaatcttcttgaacaaagagacagagaggtgtgtgaataaca---aaaccatc 2510
Dh 2529 GCCCAATTCATTTGAGAGGATCAGAGACTCAGGTTATGAGCAATGAGGCGAGCCCTG 2588
Qy 2511 actgtgtgacctcccaattagatgaatacgaactatcttccccaagtaagacct 2570
Dh 2589 CAAGTGTGACCCCTTAATATAGATGATGATGATGATGATGATGATGATGATGATGAT 2648
Qy 2571 gatcaagatatacagtc---tggttgaagcaggttccccaagccttgcgcgaagagca 2627
Dh 2649 GATGTTTCTTATGAGGTGCACATGCTGTGTGATTTTCTCCAGGCGTGGCGGAAACCGG 2708
Qy 2628 ggaagcaggttgcgaagaaggttccccaagcgttatcaacttcaagaagcagagctca 2687
Dh 2709 GGCATGGGAGCTGGAGTTCGCCAAGCTCTCTGATCATACTCTGTAAGCAACCTTATCC 2768
Qy 2688 ccagatataagtcagacatcccttgaatgaagagcctgcgaagagagagagagagagag 2747
Dh 2769 CCGCTGTCCATTAACATATACCCCATGTCAACAGAGCAAGCTGAGGATCAAGCCCAAC 2828
Qy 2748 gttcaagaatcaatcccaagagagagagagagagagagagagagagagagagagag 2807
Dh 2829 ATACAGAGACTGTTGAGACAGGAGATGATGATGATGATGATGATGATGATGATGATGAT 2888

D	5046	TTTTTGCGTGATAGAAAGCGCTTCCACACAGAAAGAAAGAACCGCCAGAGTCTTAACCAAG	5105
Q	5010	aaaatactcggaggaatattttccaagaatttgaaataccctaaagtaagatgaacacat	5069
D	5106	AAGCTACTAGAGGAGATCTTCCACAGGTTGGGATGCTCTAGATTATGGAGACTGACAT	5165
Q	5070	gctccagcttctcgttcgcacgaagtaagtcagagacgttcgaagatttgggaattgat	5129
D	5166	GGGCGCTGCGCTTCGCTCCAGAGTGATGACACAGTGGCGATCTGTGGGATGTGATTGG	5225
Q	5130	aaactgcatlgtgatalacagaccaccaagctcagagcaagtgtagaggaatgaatagaac	5189
D	5226	AAATTAACATTGTGCATACAAACCCCAAGCTCAGCGCAGGTGAAAGAAATGATATGACC	5285
Q	5190	atlaaagacccttaccaaattgacacagagactgcgcatlaatgatgtgatgtctc	5249
D	5286	ATCAGAGGAGACTTTAACTAAATTAACGCTTGGCAACGTGCTTAGAGACTGGGTGCTCTTA	5345
Q	5250	ctgcaccttgygcttlttaagtgtagaacaacccctgagacagtttgggtaccacctat	5309
D	5346	CTCCCTTAAGCCCTGTACCGAGCCCGCAACGCGCGGGGCCCATAGGCTCAACCCGATAT	5405
Q	5310	gaattgctctacgggggagacccccccggttgcagaaattgaccttgcacatagtgtgat	5369
D	5406	GAGATCTTAATATGGGGGACCCCGCCGCTTGTAAATCTCCAGACCTGACATATACAGA	5465
Q	5370	gtgtgtcttcccaagcccttgtctctctcctcctcctcctcctcctcctcctcctcctc	5429
D	5466	GTTACTTAACAGCCCTCTCTCCAAAGGTCACTTAACAGGCTCTCTACTTATGTCACACAGA	5525
Q	5430	gcgtggaagacgctccgggagagcctactcctaagagagagacttgcac-----gttccaat	5483
D	5526	GTCCTGAGACCTCTGGGGGAGACTTACCAAGAACAACTGAGACCGCACCGTGTACTCTAC	5585
Q	5484	cgcttccaaagttagagattcagctctcgttctcctcctcctcctcctcctcctcctcctc	5543
D	5586	CCTTACGAGATCGGGCAGACAGTGTGGTCCGCCGACACACAGACTTAAGAACTTGAACCT	5645
Q	5544	cggttggaaaggaccttatctcgttacttttgaccacacacaaagcgctgtgtaaagtcgaaga	5603
D	5646	CGCTGGAAGAGACCTTACACAGTCTGCTGAGACCAACCCACCGCCCTCAAGTAGAGCGC	5705
Q	5604	atcccacccgtgagctccatgcatacccaagcttgaagcgcg-----gcacact	5648
D	5706	ATCCGACTTGGATATACAGCCGCGCCACGTGAAGGCTGCGACACCCGGGGGTGAGCACTCC	5765
Q	5649	cccgactcgggtgtggaagacggaanaagacttgaaatccaccttaagcttgcgtccatcgc	5708
D	5766	TCTAGACTGACATAGCGCGCTTCAACGCTCTCAAAACCCCTTAAAAATTAAGTTAACCCG	5825
Q	5709	ctggttccttactctacaataacttcccagagcagtg-----taacgcctatagaca	5764
D	5826	GAGGCCCTTAATCCCTTAATTTCTTGATGCTCTAGAGGGGTCAAGTACTGTTCCGCCG	5885
Q	5765	gtctcgaaaccccaatagaccttatcccttaccctgctcttattatgacccgtgaagcgt	5824
D	5886	GCTCCAGTCCCTCATCAAGTCTAATATATACACTGGGAGAGTAAACAATGAGAGATC---GAG	5942
Q	5825	tcactgtaaatagcaetctgagggtgttgcctcctagagacacctgtgtgcctgaac-----	5878
D	5943	AGACGGATATGGGCAATCTTGGCAACCAACCTGTGTGACCTGTGTGGCTGACCTTAACCC	6002
Q	5879	-----tgcatttctgcctcctcgtatgattaaaccccgctttaaagagcacctcccaa	5930
D	6003	CAGATTATATGTATGTATAGCCCAACATGAGACATCTTATGGGGCTTGAATATCAATCCC	6062
Q	5931	ccctagatcc-----gtagtatagggtctatctgtcgtcccaagcacagagaaagaga	5980
D	6063	CTTTTCTCTTCCCCCGGGGCCCCCTTGTGTCTCAGGGGAGACAGCCCGAGCTGTTCGA	6122
Q	5981	aatacgtgtgggttctctgaggaatcctt-----ctgtagagatagagag	6023
D	6123	GAACTGCGCAAGACCTTAACTCTCCCTACACCCCTCGGTGCAACACTGCTGTGAACAGAC	6182

QY	6024	ctggcgctccactccaacgataaggagactggaaatgycgcgatctctctccaagaccgggtataa	608
Db	6183	TCAGCTAGACCAAGAACATCTAATATCAAAATGAGAGGATTTTATATGTTTGGCCGGGGCCCC	624
QY	6084	attctccttgcatactccggcccgccgaagtacaaaaatga tgaact-----613	
Db	6243	ACCGCCCCCGAATGCCAAGTCATGTGGGGGTCCAGACTCTCTTACTGTGGCTATTGGG	630
QY	6132	-----ataataagaagagctgcctcccacacagactagatatac	617
Db	6303	GCTGTGAGACAACCGGTAGAGCTTACTGGAAAGCCCTCATCATATGAGGATTTTCATCAAG	636
QY	6173	taagaataagttcactgacgaagaagaaacagaaataataatcaaaatgataatgtatg	623
Db	6363	TAAACACACATCTCTCCTCTGACCAAGCTGTCCAGGTATGCCAAAGATTAATATATGGTGCA	642
QY	6233	agctcggggaatagttttattatataatgycggggagagcagtgctcaacttaccatctccg	629
Db	6423	ACCCCTTATGTTATTTGGTATTACAGACGGCCGGAGACGGTACTCTCTGGACCAAGAAC	648
QY	6293	taagataagagacggggacgaacccctctgagcaatgggaaccgataaagtctgctg	635
Db	6483	ATTACTGGGCTTACGTTTGTATGTCTCCGGACAAAGATCCAGGGGCTTACATTTGGGATCC	654
QY	6353	aacaaagggcccgcccgccctgagccaccc--gcaatacttgcggtgcccacaataact	640
Db	6543	GACTCAGATACCAAAATCTAGAGCCCCGGCGCCCAATAGGGCCAAACCCGTTCTGGCAG	660
QY	6410	cgctcgcgccgcacataaacaaagccgcctcaacagatacactgatagtatctctaca	646
Db	6603	ACCAACACGCCACTCTCCAAAGCCCAACACCTTTAAGTCCCTTCAGTCCACCAACCCCA	666
QY	6470	acaagccagaaactccccc-----aggtgtcctgttaagacagaagaacacttca	652
Db	6663	GTGGAGCTCTCTCTCCCGCTTACCCCACTTCCACCGGGGAACGGAAATAGGCTGTAA	672
QY	6524	gtctcaatccagggagcttccaaagcaatcaactccacagcaaccctgatagcacttctt	658
Db	6723	ACTTAGTAGAGCGAGCCTTACCACACCCCTCAACCTCAACGACTCTCTGACAAACCCAAAGAT	678
QY	6584	gttgccttgcctactccgaaggcctctctatatataaggatagtgataagaaagaataa	664
Db	6783	GCTGTTGTGTCTAGTAGCGGGAGACCCCTTACTACGAAGGGGTGTGCTGTGGGTACT	684
QY	6644	tcaatlygaccaaagaagatagataatcaatgatacatalgaggggtlcccgcaaaataagctacc	670
Db	6843	ACTCCAAACATACCTGTGTCACAGCCAACTGCTCGTGGCTCCCAACCAAGTTGACCC	690
QY	6704	lcaactgaagttccgggaagggagacatgcatagaaagaactcccccacatcccaacaac	676
Db	6903	TCTCCGAAGTGACCGGACAGGGACTGTGCAATAGGACACTTCCCAAAACATCTAGAGCC	696
QY	6764	tttgctatgtaactgtygttllatlgagagagcccgagaataatgaatattatagactgtt	682
Db	6963	TATGTAATACCAACCCACACAAAGCAGTGCAGGCTC-----CATTAATCTAGTTGGCCCTA	701
QY	6824	ataacaggtggtggcaatgcataactggtgttaaaccctgtgtttccacactcagcttca	688
Db	7017	CAGGTACCATGTGGCGTTGTAGTACCGGGGCTTACTCATGATCTCCACACCATATACGA	707
QY	6884	accacatccaagaatttctgtgtcaatgctcaaaatgtlcccccgaatgtaatacactctg	694
Db	7077	ACCTTACCACTGATATTGTGTCTTGTGCMACTTGCCAAAGAGTCACTATCAATTCC	713
QY	6944	aggaagtgtctctgatalgaatataactatcgttatataccgacacaaagaagaacccgat	700
Db	7137	CCAGCTATGTTTACGGGCTGTTTG---AGAGATCCAAACCGACCAACAAAGAGACGGGT	719
QY	7004	cccttaccacagatgtaatgctctcgattttaggaagcgccgttggcgttagaacaaggagac	706
Db	7194	CGTTAACCTCGGCCCTATTATTTAGGTGGCTAAACCATGGGGGGAATTCCTCCCTGGAAATAG	725

Db	6200	CTCTTACTCATGACGCCCACTTGGATGACTGCAACACTTTTAGGAGACCTTGCTAGAC	614.1
Qy	1383	accgagagcgagagaaattcgttagagcgctagaaaaalgtctcgggpcgacgg	144.2
Db	6140	GGAGAAAMAAAACAGCGAGTGTCTTAGAGGCCCCCAAGCGCGGTTCGAGGGAGACGGA	608.1
Qy	1443	cgaccgcgagttgcaaaatgtagatgacatggaattcccttgaactcgcgcgggttgg	150.2
Db	6080	CGCCCAACTAGCGCCCAATGACATTAATGATGCTTTTCCCTTGGAACGTCCCAACG	602.1
Qy	1503	gactacaacacggtgaaggtaaggagagcttgaaaatctacgcaccaagctcgtgtggc	156.2
Db	6020	GACTACAAACCCACAGAGGTAGGAACACACTAGTCCACTATACGGCAAGTTGCTCTTAGC	596.1
Qy	1563	ggtctccgggpcgtcctaagaacgcccactaaatttgctaaggtaagaagtgatgcag	162.2
Db	5960	GGTCTCAAAACGCGGGCAACAGCCCAACCAATTTGGCCAAAGGTAAAGGATTAACGAC	590.1
Qy	1623	ggaccggaacgaacctccctcgtgatttctctggagagctcatgaaagccttcagcggttc	166.2
Db	5900	GGACCTTATAGTCTCCCTCAGCCCTTTTAGAGAGACTCAAGAGGCGCTTATGCGAGATAC	584.1
Qy	1683	accctcttgatccctaccagagcgccgaagaagcctaagtgcctcgcgtcctaatggg	174.2
Db	5840	ACTCTTATAGCTTGAAGACCCAGGGGCAAGAAACCAATGTGGCATGTCTATTCTGCG	578.1
Qy	1743	cagtcgctctgtgatalcagaagaactcagagactggaagggttacagagagcttag	180.2
Db	5780	CAGTCCGCCCGGATTCGGGGCCAAAGTTAGAGCGGTTAGAGATTTGAAGATAGAAC	572.1
Qy	1803	ttaacgtatcagttagagagagcgacagaaggtgttatccagaaggagacagaagagag	186.2
Db	5720	TTTAGAACACTTAGTAGAGGGAAGCTGAAAAGATCTTTAATTAACGAGAAACCCCGGACAA	566.1
Qy	1863	aaggaaacagagaaagaagagagagaaagaagaaggaagaagcgtgatagagcgaa	192.2
Db	5660	AGAGAGAACGTATTAGGAGAGAAACAGAGAAAAGGAAGAGCGCGTAAAGGACATAGAT	560.1
Qy	1923	gagaagaatttgaactaagatcttgcgcgcgaagtgtgtgaaggaaagacagcagggagga	198.2
Db	5600	GTCGAGAGAGAAAGGAGAGGACCCGACAGACATAGAGAAATAGTA-----A	555.1
Qy	1983	gagagagattttagaanaattaggtcagggccctagacagtcagga-accttggcaatag	204.1
Db	5550	GTTTCTGCTACTGTCTGTAGCGCGAGACAGACAGATATAGAGAGAGAGCGAAGGAG	549.1
Qy	2042	gaccccactcgacaagagacaggtgtcgtatgttaagaanaagacacttggcagaagaa	210.1
Db	5490	GCCCAACTGACACACACAGTGTCTTACTGCAAAAGAAAGGAGCAATTGGCTTAGAGA	543.1
Qy	2102	ctgcgcgaagaaggaaacaaagacccgaagtcctagctcagaagaag-----a	215.2
Db	5430	TTGGCCCAAGAACCAAGAGAGACCCCGGGGACCAGACCCCAAGGCTCCCTCTGACTT	537.1
Qy	2153	taaaagattaggggaagcgggtgttcggaacccctcccccgaagcccaaggttaacttgaagt	221.2
Db	5370	AGACGATTTAGGAGAGTCAAGGGTCAAGAGCCCCCTTAACCCAGATTAACCTCAGAGT	531.1
Qy	2213	ggagagggaacacagttgagttcctgtgtataccggaagcggacatlaagtctgtctaca	227.2
Db	5310	CGGGGGGCAACCCGTACCTTCTTAAGTATATCGGGGCCCAACACTCGTGTCAACCCA	525.1
Qy	2273	acasttgaagaataaagaanaaaatcccttgggtgtgctggtgtgcacagggcaacggca	233.2
Db	5250	AAATCTTGAGCCCTTAGTGACAAAGTCTGCTTGAGTCCAAAGGGGCTACTGGAAGGAGACG	519.1
Qy	2333	gtatcactgtactaccggaagaacggttgaacttggagtgaggagcgggtlaaccactggt	239.2
Db	5190	GTAATCGTGGACGACGAGTGGCGAGTGCACCTAGGCACCGGTAAAGTACCCATTTCTT	513.1
Qy	2393	tcgtgtcaatccctgagtgcccagtagtaccctcttagtagaagcttaactgtaccacaagatgg	245.2

Db	5130	CTGCATGTACCAATGGGCCCTATCTCTGCTAGGAAGAGATTGCTGACTAACTATAA	5077
Qy	2453	agctcaaatctctctttaa---acaaggaagaccagaagtgctgtgtataaacaacccat	2505
Db	5070	AGCCCCAAATTCACTTTGAAGGATCAGAGAGCTCAGGTTGTGGACCAATGGGACACCCCT	5011
Qy	2510	cactgtgttcgcccctccaattagatgtgatatctgactatattctccccaagtaagcc	2565
Db	5010	GCAAGTGTGACCTTAACATAGAGATTGATTCGGCTACATGAGACTCTCAAAAGGGCC	4955
Qy	2570	tgataaagatalatacagtc--tggttgcagcagtttccccaaagccttgcagaaacgcg	2626
Db	4950	AGATGTGCTCTTAGGGTCCACATGGCTCTCTGTATTTTCCCAAGCCTGGGAGAAACGG	4891
Qy	2627	aggagtggttgcgcaagcagaagttcccccacagttaltcaactgaagccagtgctac	2688
Db	4890	GGGGATGGGGCTGGCCGCTTGGCCAAAGTCTCTTGATCATACCTTGAAGCAACCTTAC	4833
Qy	2687	accagatcagtcgacagatcccccttgtagtagagagctgcggaagaagattgcgcga	2746
Db	4830	CCCCGTGTCAATAAACATTAACCCCATGTACAAAGAACCGACCTGGGATTAAGCCCA	4777
Qy	2747	tgattcaagattlaalccaaagggacatcctagttctcgtlccaaatcccccttggaaatcc	2800
Db	4770	CATACAGAGACTGCTGGATCATGGGAATTCTGGTACCTCGCATCCCCCTGGACACGGC	4711
Qy	2807	cctgtacacggtttagaagccttgcgaccaatgatcatcgacagatcacaggacttgaga	2866
Db	4710	CTCTGTACCCCTTAAAGAACCGGGGACTATGTATTAAGCCCTGTCCAGATCTTAGA	4651
Qy	2867	ggtcaataaagagtgacagacataaccccgaaggtccgaacctataacctttag	2922
Db	4650	AATCAACAAGCGGGTGGAGAACATCAACCCACCGCTCCCAACCTTTACACCTCTTGAG	4591
Qy	2927	cgccctccgcgctgaagcgaactgtgtacacagttatggacttlaaagaatgcctctctg	2986
Db	4590	CGGGCTCCACACGTCGCCACAGCTGTACTGTGACTTGCTTAAAGATGCTTTTTTCG	4533
Qy	2987	cctgcgattlacaccoccatagccaacacttttgccttcgaaatgtagagatccagtiac	3046
Db	4530	CTGTGACTTCCACCCCAACAGTCAGTCTCTTCCGCTTGTAGGTGAGAGATCCAGAGAT	4477
Qy	3047	gggaagaacccgggaagcttaactgtgaccgactgcgccaaaggtttaagaactcccgac	3100
Db	4470	GGGAATCTCAGAGCAATTAACCTGTGACCAACTCCGACAGGTTTAAAAAAGTCCAC	4411
Qy	3107	catctttagaagaagccctacaagagaccttgcgccaatttagatccaacaccttaagt	3166
Db	4410	CTGTATTATTAAGCCCTGCACAGGACCTTCGAGACTTCGGATTCAGCACCCAGACTT	4355
Qy	3167	gaacctctccagtcagctgtgatacctgtcttgcgcggagagccaacaacaggaactgct	3226
Db	4350	GATTCTGCTCCAGTATGTAGATGACTTACTGTGGCGGCCACTTTCGAGCTTGAGTGTCA	4291
Qy	3227	agaagtgacgaaggaactactgtctgaaattgtctgcactaggctlacagagcctctgctaa	3286
Db	4290	ACAAGTACGCGGGCCCTGTATCAAAACCTTAGGGGCACTCGGATATCCGGGCTCGGCCAA	4231
Qy	3287	gaagggccaagtttgcagaggaagaggttaacatacttgggttlaagtttgcgggcggga	3346
Db	4230	GAAGCCCAAAATTTGGCCAGAAACAGTCAAGTATCTGGGGTATCTTCTTAAAGAGGGTA	4177
Qy	3347	ggagatgcgtgcagaggaagcaagaanaactgtagtccaatatccggcccacaacagag	3400
Db	4170	GAGATGGCTGACTGAGGCCAATAAAGAGACTGTGATGTGGGACACCTTACGCCAAGACCC	4111
Qy	3407	caaacgaatgagagaatttttgggacagcttgaaatttgcagactgtgatacccgaggtt	3466
Db	4110	TGCAGACATTAAGGAGATTCTTAGGGAGCGCAGGCTTCTGCGCTCTGATCCCTGGGGTT	4051
Qy	3467	tgcgaccttagcagcccacactctacccgcgtlaacccaagaaaaagggagattctctcggc	3522
Db	4050	TGCACAATATGACCCCTTTGATACCTCTTCACCAAAACCGGACCTCTGTTTGAAGTGGG	3991

Db	3690	CCATGCAAGTAGAGGACACTAGTTAAACAACACCCTGATGCGGTGGTCCAAAGCCGAAAT	3631
QY	3887	gaaccacatcaaaagccgtctctct---cacagaagggtcaactctgtccaccagccgc	3943
Db	3630	GACCCACTACAGAGCTCTCTGCTTCTTGACACAGCCAGGATCCAGTTCCGAGCAATTAATGGCC	3571
QY	3944	ttctaacacctgcacactctctctgcctgaagagacatgataagacagtgactatgattgcca	4003
Db	3570	CTTAACCCACGACTGACCTCTCCCTCTA---CCTGAGAGGGGCTGCACATACATCTGCTT	3514
QY	4004	tcaactatgattgagagagactcggggtctccgaagaagacttacaagacataccgtctgac	4063
Db	3513	TGACATCTTGCGTGAAGGCCACGGAAGTACAGCAGATCTTACGAGCAGGCTCTCCACGA	3454
QY	4064	agaagtgtcaactcgtgttcaactgacgggaagcagctatgtgtgtgaaggtagaagggtttgc	4123
Db	3453	CGCGACCAACCTGGTATCACAGATGGGAGCGCTTCTGTGAAGAGGGGCAAGCGACAGGC	3394
QY	4124	tggggcgagctgtgtgagcgggagccgcacgactctggcgagcagcctgtccggaaagaac	4183
Db	3393	CGGAGCAGCAAGTAAACACCGAGACCGAGGTATCTGGGCTCAAGCACTGCCAGCCGGAGC	3334
QY	4184	ttcagcgaaaggctgtgactcaatgtgcctcaagcaagcttggcgcttggccgaaggaa	4243
Db	3333	ATCGGCCCAAGAGCGCTGAGTTGATGATACCGCTTCACCAACCTTAAAAATGGCAGAGGTAA	3274
QY	4244	atccataaacttatacagcagcagcagatgactcttltgcagtgcacagctacacgggc	4303
Db	3273	GAACTGAATGTTTACACCGATAGCCGTTATGCTTTTGCCACTGCGCATATTTCCAGGAGA	3214
QY	4304	catcatataacaaagggtgtgtcttaactccagcagggaggaataaagaacaaagaga	4363
Db	3213	AATATATAGAAAGGCGGGTGTGCTCATCAATCAAGAGAAAGAAATCAAAATAAGGACGA	3154
QY	4364	aattcgaactatitgaagccttaacttgcgcaaaaggcctagctatatatacacttcc	4423
Db	3153	GATCTTGGCCCTACTTGAAAGCTCTCTTCTCTGCCAAGAAAGCTTAGCATTAATTTCAATTGCC	3094
QY	4424	tggacatcagaagaacccaagaatctcalacttagagaggaaaccagatgctcagcggtgtgc	4483
Db	3093	GGGACATCAGAGAGGAAACCGCGCGGAGCAAGGGGCAACAGATGGCGGACCAAGCGGC	3034
QY	4484	caagcagcagcccaagcctgttcaactctgtcctataatagaagaagcccaagccccaga	4543
Db	3033	CGGAAACTATACCCTAGAGAAAACTCCAGAGATTCACACTTCTGATAGAAAATTCAAC	2974
QY	4544	accg-aga-cgacgtataccactagaagactctgcgaagagaataaagaataagaccagttct	4602
Db	2973	CCCCATATCTATGAACATTTTTCACATAGGGTACATGAACATTAAGAAATCTGACATAACT	2914
QY	4603	ctgagactcggaggggagc-----tgtataccctatatltgggaaggaalctcggcc	4654
Db	2913	AGGGGCCACTTATGACATGACATGCAAGAAAGTGTGGTTTATCGAGGAAAGCTGTAAATGCC	2854
QY	4655	ccacaagaaggtttagaatatgtccacaagatacactgtctcaaccacacttagaagactaa	4714
Db	2853	TCATCAATTTCAACTTTTGAACATTAATTAAGTTTCTTCATCAATTTGACCAACCTCAGTTTCTC	2794
QY	4715	acacctgcagcagatgtgtcagaaca-----tcccctatacatgtctcgaagctaccag	4768
Db	2793	AAAAACAAGGCTCTTTAGAAAGAAAGTACTCTCTCTCTTATTCATGCTTGAAACCGGATGC	2734
QY	4769	agtgtcgaactcogtgtgtcaaaacttgtgtgcctccagcctggttgaatgatactcttc	4838
Db	2733	AACGCTCAAGACATCACTGAGACTTCCCAAGCCTGTGCACAGTCAATGAATGCCAGAACTC	2674
QY	4829	cagaataccctccaaggaagaagatacgaagggaagccaccacagcgctcacttgggaagtga	4888
Db	2673	TGGCGTCAAAACAAGGACGATTAAGTTTCAAGGAGCACCGCACCGGACCTCGGAATATGA	2614
QY	4889	cttcaactaggttaagcccgcgcttaactcggaaacaaatctatcttggtttltgtagaac	4948
Db	2613	TTTTCATGAGGTAAACCTGGCTGTATGGGTATTAATTAATCTTTTAGTTTTCATGACAC	2554

QY	4949	ctttcagatggtgtagaggtctatctcactcraagaagaagcttcaacgtgtgtgttaa	5008
Db	2553		
QY	5009	gaataactgaggaatatcttccaaagatttgaa tactaagttaatagttcagacaa	5068
Db	2493		
QY	5069	tgctcagcttcglttgcccagtgaaagtcagagacttgcccaagatattggggatgttg	5128
Db	2433		
QY	5129	gaactcattgttgacatacagaccccaagctcagaagcgtatagagatgaaataaac	5188
Db	2373		
QY	5189	cattaaagagaccccttaaccaattgacacagagactgtgcattaaatgtatgtgctc	5248
Db	2313		
QY	5249	cttgcccttgttgcttttaggttgagagacaccccttgacagtttggtgcgcacccta	5308
Db	2253		
QY	5309	tgaattgctcactcggggagaccccccccglttgacagaatattgcctttgcatagtgtc	5368
Db	2193		
QY	5369	tgctgtctctcccaagccttgtctctcctaggtcgaagggctcgagtggtgtgaagcgc	5428
Db	2133		
QY	5429	agcgttgaagacagcttcgcgagagcctaactag-----gaggagacttgcaagltcoca	5482
Db	2073		
QY	5483	tcgcttccaagtttgagatcattcagttcgtatgttgaagccacgcgttgaagaaacttcgaac	5542
Db	2013		
QY	5543	tcgggtggaagagaccttacttcgttactttgacacacacaacagcgtgttgaagtcgaag	5602
Db	1953		
QY	5603	aatcccccacttggatccatgcatcccaagcttaagccgcg-----gcacc	5647
Db	1893		
QY	5648	tcaccaactcgggtgtgagagccggaagaaagacttgaatccctctaagcttcgcctccag	5707
Db	1833		
QY	5708	ccgtgtctcactccttaacaaatactcccaagccagtgatgaagccttataagcaact	5766
Db	1773		
QY	5768	cgaaagccca 5777	
Db	1713		
		ccggcttcca 1704	

RESULT 14
PCT-US93-08041-9/c
Sequence 9, Application PC/RUS9308041
GENERAL INFORMATION:
APPLICANT: The Public Health Research Institute of the City of
APPLICANT: New York, Inc.
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Davis Hoxie Faithfull and Hapgood
STREET: 445 Rockefeller Pl.

Db	3393	CGGAGCAGCA	GTAAACCCGAGACCGGAGTACTGGCCAAAGCACTGCCAGCGGAC	3334
Qy	4184	ttcagcgcaaaa	aggcttgagctcatgtgcccctcaacgaagctttgcygcctggccgaagga	4233
Db	3333	ATCCGCCCAAA	AGGCTGATGTTGATATACCGCTTACCCCAAGGCTTTAAAAATGGCAGAAAGTAA	3274
Qy	4244	atccataaacat	tataatcacgacagaggtatgcctttgagctgacacgctacagagggc	4303
Db	3273	GAACTGAATG	TTTATACCCGATAGCCGATAGCCGTTATGCTTTTGGCACTGCCATATTCCACGGAGA	3214
Qy	4304	catctataacaa	aggggtgtcttctacccacagggagggagaaataaagaacaaagga	4363
Db	3213	AATATATGA	AAAGGGCGGGTGTGCTCATCATCAGAAAGGAAAAAATCAAAAATFAGGACGA	3154
Qy	4364	aattctaaagc	tattagaagccttaccatltgcccaaaaggtagctatatatacactgtcc	4423
Db	3153	GATTTTGCC	CTACTGTGAAGGCTCTCTTCCGCCCCAAAACACTTACATATATTCATTTGCCCC	3094
Qy	4424	tggacatcaga	aaagccaaagatctcatctctgagggaaacgaatggctgacgggttgc	4483
Db	3093	GGGACATCA	GAAAGGAAACCGCGGAGGCAAGGGGCAACAGATAGCGGACCAAGCGGC	3034
Qy	4484	caagcagcgac	ccccagagctgttctacacctgtccctataatagaacgcccaagccccga	4543
Db	3033	CCGAGAGT	ATGACCCTAGAAAMTCCAGAGACTTCCACACTTTCGATGAAAAAATTCAGC	2974
Qy	4544	aacc-aga	gcacagtaacacctagaagactgycgaagagataaaaaagatagaccagttct	4602
Db	2973	CCCCATAT	CACTCATGAAATTTTCTACATACGGGATCGACATGAAAGATCTGCTAAACT	2914
Qy	4603	ctgagactc	cgagggagacc-----tgcataacctatctggagaaagaaatccctgc	4654
Db	2913	AGGGGCCA	CTTATGACGATCANAAGATGTTGGTTTTCAGGGAAACCTCTGATATCC	2854
Qy	4655	ccacaagaag	ggttgaatatltccaaagatatacgtlctaaccactagaagactaa	4714
Db	2853	TGATCAAT	CACTTGACTTAACTATTAGATTTTCTTATCATTAATGACCCACTCATGTTCTC	2794
Qy	4715	aacccgtcag	cagttgtgtcagaaca-----tcccctatcatgtctcgaagctaccag	4768
Db	2793	AAAAACA	AGGCTCTTCTTACAAAAGAACTACGTCTCTTATTCATGACGTGAACCGGATCG	2734
Qy	4769	agtgctgac	tcggtgttcaaatgtgtgcccctcgaagctgggttaatgctaatccctc	4828
Db	2733	AAGCTCA	AAAGCATCTCAGTCACTGAGCTTGCCACAGCTTGTGCACAGGTCAATGCCAGCAATC	2674
Qy	4829	cagaatac	ctccagaagaagactaaagggaaagccacccagagcgctcatctggaagtga	4888
Db	2673	TGGCGTCA	AAACAAAGGAGCTAGTGTGAGGGGACCGCACCGGACCCACTGGGAAATTGA	2614
Qy	4889	cttcactga	aggtlaaagccggtctaataacggaacaatactatgtgttttgttagcac	4948
Db	2613	TTTCACTG	AGGTTAAACCTGCGCTGTATGAGGTATTAATCTTTTATGTTTATGTAGCAC	2554
Qy	4949	ctttcagga	tatagggtlaagagctctatctctactaagaagaagacttcaaaccgtgtgtctaa	5008
Db	2553	TTTCTCTG	AGGTTAAACCTTTCACAAAGAAAGAACTGCCAANAAGTTGTATCCAA	2494
Qy	5009	gaaatactg	gaagaaatlttccaaagatttgaaataccaaaggttaataggttcagacaa	5068
Db	2493	GAGGCTAC	TAAAGAAATTTTCCCGAGATTTCGGCATGCCACAGATATTGGAAACGACAA	2434
Qy	5069	tgtctcag	cttctgtgtccaggttaagtcagggaactgtgcccgaagatatvgggatgtattg	5128
Db	2433	TGGGCTGC	CTTCTCTTCCAAAGTAAGTACAGACATGACGATGCCGATTAATTATGGGGTTGATTG	2374
Qy	5129	gaactgat	tgtgtcatcagagcccaagctcaggacaggttagaggaagatgaatagac	5188
Db	2373	GAAACTAT	ATTGTTCTTACAGACCCACAGATTTAGGTACAGGTAAAGATGAATATGAGAC	2314
Qy	5189	cattaaaga	gccttaccaatgtgacacagagactgtgcatatgaattgttagtgcct	5248
Db	2313	AATCAAGA	CACTTAAATTGACGCTTGCAACTGCGCTAGGAACTGGGCTCTCT	2254

OY	5249	ccgcgccttgctgtctttagaggtgaggaaccccttgaaagaggttgggcctgaccccta	5308
Db	2253	GCCTTCCCTTGCCCTGTATGGAACCCGCAACAGCGGGGCCCCACGGCTTCACCCCATTA	2194
OY	5309	tgaattgcttaagggaggaaccccccgcttgacagaatltgcttgcacatagtctga	5368
Db	2193	TGAATATCTTATATGAGGGACACCCCGGCCCTGTGAACCTTCCTGTATCCGTGACATGGCAAA	2134
OY	5369	tgtgcgtcttcccaagccttltctcttagagtcctaaagcgctcgatggtgtgagcgagc	5428
Db	2133	GGTTATCCATAAACCCCTCTCTCCACAGCTCATTTACAGGGCACTCTACGTGTCACACAGA	2074
OY	5429	agcgltgaaagcagctccggaggagcctctacag-----gaggagacttgcaagttcacaa	5482
Db	2073	AGCTGTGAAGACCGTTGGCGGCACCTTACCAAGACAACTAGACCGCGGTGATGACTCTCA	2014
OY	5483	tgcgttccaagtgtgagatcaagtlcaltgltagaagccacgctgacggaanaacctcgagc	5542
Db	2013	CCCTTTCCGGGTGGGTGACACAGTGTGGGTCCGACAGCACCAACTTAATAATCTGAGAAC	1954
OY	5543	tcgatgtgaagggagacctatctcgtactttgacacacaaagcgtgtgtgaaagtctgaag	5602
Db	1953	CCGCTGGAAGAGACCCCTATACCGCTCTACTGACTGATACCCACCGCTCTCAAGATGACGG	1894
OY	5603	aatcccaacctgatatcatatcatcccaagtlaagccgcg-----gcacc	5647
Db	1893	CATTGCACACGTGATTCACACCTCTCCACGTAAGAAAGCTGCCGACACACAGATTTGACCCAC	1834
OY	5648	tcccgcactcggggtgtgagagcggaaaaagctgaaatccctctaagcttcgcctccatcg	5707
Db	1833	AGCAGAAATGCAGATGGCGGTTCACAGCTCCCAAAATCCCTTAAGATTAAGTTGACCCG	1774
OY	5708	cctgttcttactacttaacaatactcccgagcgagtagtaaaagccttatagacagct	5767
Db	1773	CGGAGACCTCCTAATCCCTTAATTTCTTCTCTCTCTCAAAAGGGGCGAGATCGCAGCA	1714
OY	5768	cgaaccccga	5777
Db	1713	CCGGGCTCCA	1704
RESULT	15		
US-09-111-085-1			
Sequence 1, Application US/09111085			
Patent No. 6100034			
GENERAL INFORMATION:			
APPLICANT: Stoye, Jonathan P			
APPLICANT: Weiss, Rodin A			
TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope			
TITLE OF INVENTION: specific sequences			
FILE REFERENCE: 4238/75168			
CURRENT APPLICATION NUMBER: US/09/111,085			
CURRENT FILING DATE: 1998-07-07			
EARLIER APPLICATION NUMBER: GB 9710154.7			
EARLIER FILING DATE: 1997-05-16			
NUMBER OF SEQ ID NOS: 16			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 1			
LENGTH: 2462			
TYPE: DNA			
ORGANISM: Porcine retrovirus			
US-09-111-085-1			

Query Match	19.1%	Score 1568.4	DB 3	Length 2462
Best Local Similarity	79.6%	Pred. NO. 0		
Matches 2002: Conservative	0	Mismatches 396	Indels 116	Gaps 8
QY	5410	tcgaatggatgacgcagcgcctgtgaagaacctccggagagcctacttaagaaggaact	5469	
Db	1	tcgaatggatgacgcagcgcctgtgaagaacctccggagagcctacttaagaaggaact	60	

QY	5470	tgcaagttccacatcgcgtcccaagttgagatctagctatglttagacgcacgcgtgcaag	5525
Db	61	tgcaggttccacatcgcgtcccaagttgagatctagctatglttagacgcacgcgtgcaag	120
QY	5530	gaaacctcgagactcgcgtggaagagacctatctcgtacttggaccacacaaagcgtcg	5585
Db	121	gaaacctcgagactcgcgtggaagagacctatctcgtacttggaccacacaaagcgtcg	180
QY	5590	tgaagctgcaagagatccccccacgtgatactgcatccacgttgaagccgcgcaccc	5645
Db	181	tgaagctgcaagagatccccccacgtgatactgcatccacgttgaagccgcgcaccc	240
QY	5650	ccgactcgggtgtgagagccgaaaagctgagatcccccttaagcttcgcctccatcgcc	5705
Db	241	ccgactcgggtgtgagagccgaaaagctgagatcccccttaagcttcgcctccatcgcc	300
QY	5710	tgattccctactctcaaaataactccccagccagttagttaaagccttatagaacgcgcg	5765
Db	301	tgattccctactctctcaaaataactccccagccagttagttaaagccttvtgacagcccg	360
QY	5770	aaccccataagaccttatacccttaacctggtgtattatgacccctgatacgggtgtcaat	5825
Db	361	aaccccataagacctatctctcaacctggtgtattatgacccctgatacgggtgtcaat	420
QY	5830	gtaaatagactcagagvtgtgtcccttagagcaccctgtgtgacctgaactgtattctgc	5885
Db	421	atlaacagcactccaagggagagcttccctgtggagacctgtgtgacctgaattatgtctgc	480
QY	5890	ctccgattgattaaaccccgctgtta-----aaagacaacctcccaacctagtcctgagt	5943.3
Db	481	cttcgatacagaataacccctgtgtctcaatgacacagccacaccccccgtatctccgctgt	540
QY	5944	tatgtgtcttaattgtcccccag---cacagagaaagaaataactgtgtggtgtcttgg	6000
Db	541	taaggttttaagtttgcgcagagaccccccacaaataagaaatatgtgtggaatccctgaag	600
QY	6001	gaatcctctgttagagagatgagactgtgcctacccccaacgatgtagactgtgaaatgvcg	6060
Db	601	gattctctttgcaagcaatgtagactgtcatacttccaatgatactggaattgtgaaatgvcga	660
QY	6061	atcctctccaagggccgggtgaaaattctctctgtgtaatcc-----	6102
Db	661	gtccctcaagacagacagatgaagttaactcttgtttaacaatccctacagttataatcaa	720
QY	6103	-----ggcccggcgaagtacaatatgaataact	6131
Db	721	ttatattatgccaatgagatgtagaagattgtgcaacagcgggtgtacaataaagatvt-acg	779
QY	6132	ataataagataagagactgtcctcccatcagacttagattaacctaaagataagtttcaacty-	6190
Db	780	aaataagcaataatagctgtcatctcgtttagacctagtttaacttaaaataagatttcaactga	839
QY	6191	-aaagggaaaacagaaaatatctcaaaaagtgtgataaattgtgtatgagctgtggaataagtttt	6249
Db	840	aaaagggaaaacagaaaatatctcaaaaagtgtgataaattgtgtatgagctgtggaataagtttga	899
QY	6250	ttac-----tatatgvcggggagcaggtgtccactttaacatctgccttaagatagagac	6305
Db	900	ctatgtagagctctcgtggaagaaagagactcgttcttgactatctgcctcagaaatagaaac	959
QY	6306	ggggagacaaacccctgtgtgcaatgtgagaccgataaagttactgtgtgcaacggggccccc	6365
Db	960	tcagatgtgaacctcccggttgcctatagaccataaagttgttgccgacaagaagactcc	1019
QY	6366	ggccctgtgagtcacgcgataacttgcgcggtgtcccaatlaacctgtcgtgcgctgaat	6425
Db	1020	aatccaagaaca-----gaagccatctcc	1043
QY	6426	aaacagacgcctcagcaacagttacacgtgattgttctcttccaaacagcgtctagaactc	6485
Db	1044	taaacctctcgtattacaatacaaacctctcgtgatactgtcccaactga-----	1088
6486	ccaagtgctctgttlaagacagagaaagactcttcagcttccatccagggagcttcca	6545	

Db	1089	gacctaacatcactattaaacaagggcgaaaacttttagcctcatccagggactttta	1144
QY	6546	agccatcaactccacgcgacctgahgcacttctctgttgctgtctatccctcagg	6605
Db	1149	agctcttaactccacgactccagagcgctactcctctctgtctgtcttgcttcggtc	1208
QY	6606	gacctctatctatgtgggagtgctgtgaagaagaatttcaatgatgtacaaagagcatag	6665
Db	1209	cccccttactctatgtgggaaatggtctgtaggagggaaatctcaatgtgacaaaggaactat	1266
QY	6666	aaatcaatgtacatgtggggtgccggaataagctctaacctcagaaagttccgggaag	6725
Db	1269	agaccaatgtacatgtgggtatcccaataaagattacccttactgaggtttctcggaag	1322
QY	6726	gacatgcatatgaaaagctctccccaatccacacacaccttgcctatagtactgtgtta	6785
Db	1329	caactgcctatggagatggtgtcccccatacccaacacaccttgttaacacacgtgaagcctt	1388
QY	6786	tgaagcagcgctacgaanaactagatttagtaccgtgttaaaaggtgtgtggcatgtcaa	6845
Db	1389	taatcgaaacctctgtagagtgcaatatctcgttaactcgtgttatgacagatgtgtgtggcatgtcaa	1448
QY	6846	tactgtggttaacacctgtgtgtttccacctagttcttcaacccaatccaaagaattctgtgt	6905
Db	1449	tactgtggttaacacctgtgtgtttccacctgtgttttcaacccaatgaagacttttgcgt	1508
QY	6906	catgtgtccaatctgtcccccagagtgactacacctctcgtggaggtgtgtcttgtagta	6965
Db	1509	tatgtgtccaatctgtcccccagagtgactactatcccgaaaagcgatcccttgtagta	1568
QY	6966	tgactatcgtataacgcgacacaaaagagaaacccgtatacccttaccctagctgttaatgt	7025
Db	1569	tgactatagatataatcgtcccaaaaagagcccatatcccttgacactagctgttaatgt	1628
QY	7026	cggtatttaaggagcgccgtgtgtgtctgtagaaacagggagacgtgtgccctgatacagaagca	7085
Db	1629	cggtatttggagtggtctgtccaggtcgtgggaacaaagaaacgtgtgcccttaatacagaagccgca	1688
QY	7086	gaaagctagaagaagaagactgtgtgtgactcaatgtgcgcatacgacaagaagatcccgagcctt	7145
Db	1689	aaagctgtgagaagaagacttgaaactacactcgaattgttaacggaaagatctccaaagcctt	1748
QY	7146	aaaagagtcgtctagcaacctagaagaagtcctcgtactcttgtctgaaagtgtgtctaca	7205
Db	1749	agaaaaactctgtcagtaacctgtagaagaaatccctcaacctcttactcgtgaagtgtgtctaca	1808
QY	7206	gaaacggggggagattgaatctgtctgttttcaagaagaaggtgtgttatgtgcagccttaaa	7265
Db	1809	gaaacggaaaggggtctagatctgttatacttccaaagaagaaggtgttatgtgtgaccttaaa	1868
QY	7266	agaagaatgtgtgtctctcaatgtagaatcactagaagagcaacaaagactccatgaanaagct	7325
Db	1869	agaagaatgtgtgtctctcaatgtagaatcactagaagagcaacaaagactccatgaagaagct	1928
QY	7326	tagaaaaaagtctagagaagcgctcgaaagggaagaagagctcgacaaggggtgtgttgaag	7385
Db	1929	tagaaaaaagtctagagaagcgctcgaaagggaagaagagctcgacaaggggtgtgttgaag	1988
QY	7386	atggttcaacaggtctcctctgtgatacgcacacctgtcttctgtctcgaaaggggcccccaagt	7445
Db	1989	atggttcaacaggtctcctctgtgatacgcacacctgtcttctgtctcgaaaggggcccccaagt	2048
QY	7446	agtcctgtctccgttacttaacagtctvggcctgtgttaattaaagttgttgtcctgtt	7505
Db	2049	agtcctgtctccgttacttaccagtctvggcctgtgttaattaaagttgttgtcctgtt	2108
QY	7506	tagaagaacgggtagtgccaatccaaagtaatgttacttgggaacaaagtaaccaagcctctc	7565
Db	2109	tagaagaacgggtagtgccaatccaaagtaatgttacttgggaacaaagtaaccaagcctctc	2168
QY	7566	gagcgaaggaacatgacctctagacttcccaagttcttlaagattgaacattcaacaga	7625

Db 2169 gagccaagagaaactgacctagcccttccagtttlaagattagaactattaacaga 2228
Qy 7626 caagaagtgggaatgaagaatgaagaatgaacccctaccgaaggaagt 7685
Db 2229 caagaagtgggaatgaagaatgaagaatgaacccctaccgaaggaagt 2288
Qy 7686 aataaaagctctaataatgccccgaattacagaccctgtgtgctgccaagtaaatagtag 7745
Db 2289 aataaaagctctaataatgccccgaattacagaccctgtgtgctgccaagtaaatagtag 2348
Qy 7746 aaggtcacacttcctattgttccaggccctgtatcccttgccctaaagtaagatacagaa 7805
Db 2349 aaggtcacacttcctattgttccaggccctgtatcccttgccctaaagtaagatacagaa 2408
Qy 7806 atgagtgactaatgcttattgtgattctgtaaaactgactgacacataga 7859
Db 2409 atgagtgactaatgcttattgtgattctgtaaaactgactgacacataga 2462

Search completed: February 24, 2002, 02:58:34
Job time: 21265 sec

OM of: US-09-171-553B-4 to: Issued_Patents_NA.* out_format : pfs
Date: Feb 24, 2002 10:12 AM

About: Results were produced by the Gencope software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame-pzn.model -DEV=x11
-O=cg2_1/USPTO.spool/US09171553/runat.22022002.122826.11171/app-query.fasta.1.3312
-DB=Issued_Patents_NA -OPMT=fasted -SUFFIX=feb22std.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPTC=0.000
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -TGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blotsum62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09171553.@CGN1.1.158 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPX -WAIT -THREADS=1

Search information block:

Query: US-09-171-553B-4
Query length: 524
Database: Issued_Patents_NA.*
Database sequences: 351203
Database length: 11338999
Search time (sec): 134.760000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-766-528-2	2758.00	3838.83	7.2e-206	7333	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-766-528-3	2640.00	3671.81	7.2e-197	8132	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-766-528-1	2634.00	3663.48	2.1e-196	8060	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-09-075-272-1	1664.00	2327.30	5.5e-122	8655	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-716-511-1	1660.50	2294.43	3.8e-120	8535	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-09-011-745-3	1355.50	1867.43	2.3e-96	7308	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-09-011-745-4	1355.50	1867.43	2.3e-96	7308	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-09-011-745-2	1355.50	1866.97	2.4e-96	7616	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-550-961-1	1355.50	1865.97	2.4e-96	8332	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-258-420-13	1354.50	1864.74	3.2e-96	8202	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-110-300A-8	1350.50	1858.95	6.8e-96	8323	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-866-642-8	1350.50	1858.95	6.8e-96	8323	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-110-300A-9	1350.50	1856.50	9.3e-96	10367	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-866-642-9	1350.50	1856.50	9.3e-96	10367	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-866-642-10	1344.00	1858.92	6.8e-96	3674	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-105-483-324	1344.00	1858.92	6.8e-96	3674	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-458-101-324	1344.00	1858.92	6.8e-96	3674	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-181-335B-3	822.00	1142.31	5.6e-56	789	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-181-335B-6	822.00	1142.31	5.6e-56	789	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-181-335B-5	819.00	1138.09	9.6e-56	789	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-181-335B-4	819.00	1138.09	9.6e-56	789	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-181-335B-3	816.00	1133.87	1.7e-55	789	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-181-335B-2	816.00	1133.87	1.7e-55	789	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-181-335B-1	758.50	1026.71	1.5e-49	9387	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-532-814-1	640.00	858.88	3.4e-40	9387	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-991-789A-227	636.00	858.88	3.4e-40	9387	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-991-789A-227	636.00	858.88	3.4e-40	9387	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-716-511A-3	560.00	746.11	2.6e-33	5865	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-716-511A-3	560.00	746.11	2.6e-33	5865	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-5	421.50	555.99	2.8e-23	6505	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-5	421.50	555.99	2.8e-23	6505	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-6	413.50	540.58	1.8e-22	9318	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-6	413.50	540.58	1.8e-22	9318	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-07-753-520B-4	411.00	538.34	2.5e-22	8316	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-07-753-520B-4	411.00	538.34	2.5e-22	8316	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-352-990-1	406.00	534.29	4.1e-22	6365	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-352-990-1	406.00	534.29	4.1e-22	6365	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-3	404.00	533.53	4.5e-22	5292	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-3	404.00	533.53	4.5e-22	5292	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-2	404.00	533.53	4.5e-22	5292	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-2	404.00	533.53	4.5e-22	5292	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-1	404.00	533.47	4.6e-22	5323	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-1	404.00	533.47	4.6e-22	5323	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-4	404.00	533.38	4.6e-22	5363	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-4	404.00	533.38	4.6e-22	5363	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-5	403.00	530.46	6.7e-22	6145	
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-793-610-5	403.00	530.46	6.7e-22	6145	

seq_name: /cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-766-528-2
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-935-312-1 + 403.00 530.46 6.7e-22 6145
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-848-760B-1 + 403.00 530.46 6.7e-22 6145
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-789-333F-58 + 402.00 531.47 5.9e-22 49
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-133-944-1 + 401.50 522.82 1.8e-21 1010
/cg2_6/ptodata/2/ina/6B.COMB.seq:US-08-182-612B-3 + 399.50 556.18 2.5e-23 393

seq_documentation_block:

Sequence 2, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 7333 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-766-528-2

alignment_scores:

Quality: 2758.00 Length: 525
Ratio: 5.263 Gaps: 1
Percent Similarity: 99.810 Percent Identity: 99.619

alignment_block:

US-09-171-553B-4 x US-08-766-528-2 ..

Align seg 1/1 to: US-08-766-528-2 from: 1 to: 7333

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuSerThrThr 17
|||||
598 ATGGGACAGACAGTACTACCCCTGATTGACTCTGACCATTTGAC 647
17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
|||||
648 TGAAGTTAATCCAGGCGCCATATTTTCATTTGACGTTAGAGAGGAC 697
34 rGTPolnThrPheCysAlaSerGluTrpProThrPheAspValIdlyTrp 50
|||||
698 CTGGGAGACTTTCTGTGCTGTAATGAGCAACATTCGATGTGATGAG 747

```

51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIle 67
748 CCATCGAGAGGACCTTAAATTATTCGAAATATTCCTGGCTTAAAGCCAAAT 797
67 eilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
798 CATTTTACAGACTGAGACCGGCTCTCATCTCATCAGAGAGCCCTATATCC 847
84 eutThrTrpGlnAspLeuAlaGluAspProProTyrValLysProTyr 100
848 TTACGGGCAAGATTTGGCAGAAAGATCCCTCCGCAATGGGTAAACCATG 897
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGlu 117
898 CTAATAAACAAGCAAGAGCCAGCTCCGCAATCTGGCTCTGGAGAA 947
117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuPro 134
948 AAACAACACTCGCCGAAAGTGCAGCCCTC.TCCGTATATCTACCC 996
134 rAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
997 GAGATCGAGAGACCGCCGACTGGCCGAGAACCCGCTTCCCCACC 1046
151 ProLeuSerSerThrGlyCysGlyGluGlyThr.SerAlaProProGly 167
1047 CCTTATCAGACACAGGAGTGTGAGGAGACCTCTGCTCCCTCTGAG 1096
167 lProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
1097 CTCGGTGTGAGAGGACCTCTGCGGAGCTCGAGCCGAGAGAGGCC 1146
184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyr 200
1147 ACCCGGAGCGGACAGAGAGATGCCATATTACCGCTGCGACCTATGG 1196
200 yProBrometProGlyGlyGlnLeuGlnProLeuGlnTyrTrpPro 217
1197 CCTCCCATGCGAGGGGCCAATTGACGCCCTCAGATATGGCCCTTT 1246
217 eSerAlaAspLeuTyrAspTrpLysThrAsnHisProProPheSerGlu 233
1247 CTTTGGCAGATCTCTATATATGGAACCTAACCTCCCTCTCTCGAG 1296
234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHis 250
1297 GATCCCAACAGCCCTCAGGGGTGGTGAAGCTTATGTTCTCTACCA 1346
250 nProThrTrpAspAspCysGlnGlnLeuGlnThrLeuPheThrThrg 267
1347 GCTTACTTGGATGATGTCACAGCTGTCGACAGACCTCTCACAACCG 1396
267 lGluAlaArgLysIleLeuLeuGlnAlaArgLysAsnValProGlyAla 283
1397 AGGAGCGAGAGAGATCTCTTGAAGGCTTAAATAAATGTTCTCTGG 1446
284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPhePro 300
1447 GAGCGGCGACCCAGCAGTGTGCAAAATGAGATGACATGGGATTTCC 1496
300 uThrArgProGlyTyrPaspTyrAsnThrAlaGluGlyArgGluSerLeu 317
1497 GACTCGCCCGGTTGGGACTACAAACCGCTGAAGTAGGAGAGAGCTTGA 1546
317 yIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
1547 AATTCATATGCCAGGCTCTGTTGGGGTCTCGGGGGCCCTCAAGACG 1596
334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGlu 350
1597 CCACATAATTTGGCTAAGGTAAAGAGGTGATCCAGAGGACCGACCAAC 1646
350 oProSerValPheLeuGlnArgLeuMetGluAlaPheArgArgPheThr 367

```

```

1647 TCCTCGGTATTTCTTGAGAGGCTCATGAGACCTTCAGGGGTTCACCC 1696
367 ropPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe 383
1697 CTTTGGATCTTACCTCAGAGGCCCGAAGAGCTCAGTGGCCCTGCTTC 1746
384 lIleGlyIleSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 400
1747 ATTGGCAGTGGCTCTGGATATCAGAGAGAACTTCAGAGACTGGAAG 1796
400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluValT 417
1797 GTTACAGAGAGGCTGATTAGCTGATCTAGTGAAGAGAGAGAGAGTGT 1846
417 yTyrArgArgGluThrGluGluGluGluGluGluGluGluGluGlu 433
1847 ATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1896
434 ArgGluGluArgGluGluArgArgArgArgArgGluGluGluGluGlu 450
1897 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1946
450 rLysIleLeuAlaAlaValAlaGluGlyLysSerSerArgGluArgGlu 467
1947 TAAGATCTTGGCCGCACTGTTGAAGGAGAGAGAGAGAGAGAGAGAG 1996
467 rAspPheArgLysIleArgSerGlyProArgIleSerGlyAsnLeuGly 483
1997 GAGATTTAGAGAAATTAGCTCAGGCCCTTAGACAGTCAAGAGAACTT 2046
484 AsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLys 500
2047 AATAGAGACCCCACTGCACAAAGACAGTGTGCTATTGTAAGAAAAAG 2096
500 yHisTrpAlaArgAsnCysProLysGlyAsnLysGlyProLysValL 517
2097 ACCTGGGCGAGAGACTGCCCCCAAGAGGAGAAACAAAGAGAGAGAG 2146
517 eAlaLeuGluGluAspLysAsp 524
2147 TAGCTTAGAGAGAGATTAAGAT 2169

```

seq_name: /cgn2_6/ptodata/2/ina/6b.COMB.seq:US-08-766-528-3

seq_documentation_block:

Sequence 3, Application US/08766528

Patent No. 6190861

GENERAL INFORMATION:

APPLICANT: Jay A. Fishman

TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766,528

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/572,645

FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers

REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: MGP-038CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)227-5941
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8132 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-766-528-3

alignment_scores:
 Quality: 2640.00 Length: 525
 Ratio: 5.116 Gaps: 1
 Percent Similarity: 98.286 Percent Identity: 94.476

alignment_block:
 US-09-171-553B-4 x US-08-766-528-3 ..

Align seg 1/1 to: US-08-766-528-3 from: 1 to: 8132

1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpTh 17
 585 ATGGGACAGACGGTGCAGACCCCTCTAGTTGACTCTCGACCATTTGGAC 634
 17 rgluValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
 635 TGAAGTTAAATCCAGGGCTCATTAATTGTCAGTTCAAGTTAAGAAAGGAC 684
 34 rotrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
 685 CTGGGAGACTTCTGTCTCTGATGAGCGGACGACATTCGATGTGGATGG 734
 51 ProSerGluGlyThrPheAsnSerGluIleIleuAlaValLysAlaI 67
 735 CCATTCAGAGGGACCTTTAATCTGAGATTATCTGGCTGTAAACAGACT 784
 67 eilePheGlnThrGlyProGlySerHisProAspGlnIleProtyrIleL 84
 785 TATTTTTCAGACTGGAACCGGCTCTCATCCGATTCAGAGCCCTATATCC 834
 84 eutHrTrpGlnAspLeuAlaGluAspProProtrPyrValLysProtrP 100
 835 TTACGTGGCAAGATTGGCAGAGGATCTCCGCCATGGGTTAAACCATGG 884
 101 LeuAsnLysProAspGlySerProGlyProArgIleLeuAlaLeuGlyGly 117
 885 CTGAATAGCCCAAGAAAGCCAGGCCCCGGAATCTGCTCTTGGAGAGAA 934
 935 AAACAAACACTCGGCTGAAAGAGCAAGCCCTC.TCCTCATATCTACCCC 983
 117 SASnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
 935 AAACAAACACTCGGCTGAAAGAGCAAGCCCTC.TCCTCATATCTACCCC 983
 134 rgsAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
 984 GAGATTTAGAGGAGCCCGGCTTGGCCGGAACCCCAATCTGTCCCCACCC 1033
 151 ProLeuSerSerThrGlyCysGlyGluGlyThr.SerAlaProProGlyA 167
 1034 CCTTATCTGCGACAGGGTGGCCGCGAGGGGACCTTGGCCCTCTGGAG 1083
 167 laproValAlaGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
 1084 CTCCGCGCGTGGAGGAGCTCTGCTCAGGAGCTCGAGCCGGAAGGGGCGCC 1133
 184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgTrpTyrG 200
 1134 ACCCGGAGCGGACGACGAGATTCGCGACATTACCGCTGCGCAGAGTACGG 1183

200 yProPromeLeuProGlyGlyGlnLeuGlnProLeuGlnIleTyrTrpProPheS 217
 1184 CCTCCACACCGGGGGGCCCAATTCGACCCCTCCAGTATTTGGCCCTTTT 1233
 217 eISerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
 1234 CTCTCGAGATCTCTATTAATTGGAAAACTAACATCCCTCTTCTCGGAG 1283
 234 AspProGlnArgLeuThrGlyLeuValGlnSerLeuMetPheSerHisG 250
 1284 GATCCCAACGCGCTCAGGGGCTTGTGATGATCCCTATGTTCTCTACCA 1333
 250 nProThrTrpAspAspCysGlnGlnLeuGlnIleThrLeuPheThrThrg 267
 1334 GCTTACTGGGATGATTTGTCACAGCTGCTGACAGACTCTTCACACCG 1383
 267 lugiuarArgGluArgIleLeuLeuGlnValArgLysAsnValProGlyAla 283
 1384 AGGAGCGAGAGAGAAATCTATAGAGGCTAGAAAAATGTTCTCTGGGGCC 1433
 284 AspGlyArgProThrGlnLeuGlnAsnGluIleaspMetGlyPheProle 300
 1434 GACGGGCGACCCACCGGCTTCCAAATGAGATTGACATGGGATTTCCCTT 1483
 300 uThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGlySerLeuL 317
 1484 AACTCGCCCGGTTGGGACTACACACGCGTGAAGAGTAGGAGAGACTTGA 1533
 317 ySileTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
 1534 AAATCTATCGCCAGGCTCTGGTGGGGGCTCTCCGGGGCGCTCAAGACGG 1583
 334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPr 350
 1584 CCACATATTTGGCTAAGTAAGTAAGTATGACAGGACCGAATGTAACC 1633
 350 oProSerValPheLeuGlnArgLeuMetGluAlaPheArgArgPheThrP 367
 1634 CCCCTCTGTTTTCTTTCAGAGGCTCTTGAAGCTTTCAGGGCGGTACACC 1683
 367 roPheAspProThrSerGlnAlaGlnLysAlaSerValAlaLeuAlaPhe 383
 1684 CTTTTCATCCCACTCAGAGGCCCAAAAGCTCAGTGGCTTGGCCCTTT 1733
 384 lIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlnG 400
 1734 ATAGGACAGTCAAGCTTGGATTTAGAAAAGAAAGCTTTCAGAGACTGGAAG 1783
 400 yLeuGlnGlnAlaGlnLeuArgAspLeuValArgGluAlaGlnLysValT 417
 1784 GTTACAGAGGCTGAGTTACGTGATCTAGTGAAGAGGACGAGAAAGTAT 1833
 417 yTyrTrpArgArgGlnThrGlnGlnGlnLysGlnLysGlnLysGlnLys 433
 1834 ATTACAAAGGAGACAGAGAAAGAAAGGAAAGGAAAGGAAAGAGAGAA 1883
 434 ArgGluGlnArgGlnGlnArgAspArgAspArgGlnGlnLysAsnLeuTh 450
 1884 ACAGAGAAAGGAGGAGAAAGACGTATTAACGCAAGACAGAAATAATTGAC 1933
 450 rLysIleLeuAlaAlaValAlaGlnLysSerSerArgGluArgGluA 467
 1934 TTAGATCTTGGCTGCGAGTGTCAAGGAAAGCAATACGAAAGAGAGA 1983
 467 rgsAspPheArgLysIleArgSerGlyProArgGlnSerGlyLysLeuGly 483
 1984 GAGATTTTAGGAAATTAAGTACGCGCTAGACAGTACAGGAAACCTCGGCC 2033
 484 AsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGlnLysG 500
 2034 AATAGGACCCCACTCGACAGGACCAATGTGCTATTTGTAAGAAAGAGG 2083
 500 yHisTrpAlaArgAsnCysProLysGlyLysGlnLysGlyProLysValL 517

|||||
 2084 ACACCTGGCGAAGAACTGCCCCAAGAGGAGAAACAAAGACCAAGATCC 2133
 517 euAlaLeuGluGluAspLysASP 524
 |||||
 2134 TAGCTCTAGAACAGATTAAGAT 2156

seq_name: /cgn2_6/plodata/2/ina/bh_comb.seq:us-08-766-528-1

seq_documentation_block:

Sequence 1, Application US/08766528
 Patent No. 6190861
 GENERAL INFORMATION:
 APPLICANT: Jay A. Fishman
 TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
 NUMBER OF INVENTIONS: AND METHODS OF USE
 NUMBER OF SEQUENCES: 74
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD, LLP
 STREET: 60 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/766,528
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/572,645
 FILING DATE: 14-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Louis Myers
 REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: MGP-038CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)227-5941
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8060 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-766-528-1

alignment_scores:

Quality: 2634.00 Length: 525
 Ratio: 5.105 Gaps: 1
 Percent Similarity: 98.286 Percent Identity: 94.286

alignment_block:

US-09-171-553b-4 x US-08-766-528-1 ..

Align seg 1/1 to: US-08-766-528-1 from: 1 to: 8060

1 MetGlyGlnThrValThrThrProLeuSerLeuPheLeuAspHisTrpTh 17
 |||||
 3112 ATGGGAGACAGCGTACGACCCCTTGTAGTTGACTCTCGACCAATTGGAC 3161
 17 rGlValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
 |||||
 3162 TCAAGTTAAATCCAGGCTCATTAATTTCAGTTTCAGTTAAGAGGAC 3211
 34 rGTrGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
 |||||
 3212 CTTGGCAGACTTCTGTGTCTGTGAATGGCCGACATTCGATGTTGGATGG 3261

51 ProSerGluGlyThrPheAsnSerGluLeuLeuAlaValLysAlaIle 67
 |||||
 3262 CCATCAGAGGGGACCTTAATTCTGAGATTAATCTGCGCTGTTAAGCAAT 3311
 67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
 |||||
 3312 TATTTTACAGACTGACCCGGCTCTCATCCGATCCAGACGCCCTAATATCC 3361
 84 euThrTrpGlnAspLeuAlaGluAspProProTyrValLysProTrp 100
 |||||
 3362 TTACGTGCAGAGATTTGGCAGAGATCTCCGCAATGCTTAACATATG 3411
 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyLys 117
 |||||
 3412 CTGATTAAGCCCAAGAAAGCCAGGTCGCCGAATCTGCTTGGAGAGAA 3461
 117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuPro 134
 |||||
 3462 AAACAACACTCGGCTGAAAAAGTCAAGCCCTC.TCCTCATATCTACCC 3510
 134 rGAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
 |||||
 3511 GAGATTGAGAACACCGCCTTGCGGAAACCCCAATGCTGTCCGCCAC 3560
 151 ProLeuSerThrThrGlyCysGlyGlyThr.SerAlaProProGly 167
 |||||
 3561 CCCATTATCTGGACAGGAGTCCCGCAGGGGACCTTTGGCCCTCCGAG 3610
 167 LaProValAlaGluGlyProAlaAlaGlyThrArgSerArgGlyAla 183
 |||||
 3611 CTCGGCGGTGGAGGACCTTCTGCAAGGAGCTCGAGCGGAGGGGCGCC 3660
 184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrGly 200
 |||||
 3661 ACCCGGAGCGGACAGACGATCGGACATTAACGCTCGCACCTACGG 3710
 200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPhe 217
 |||||
 3711 CCTCCACACCGGGGGGCCAATTTGCAGCCCTCCAGATTAATGGCCCTTT 3760
 217 eSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
 |||||
 3761 CTTCTGCAGATCTCTAATTAATGAAACTAATACCATTCCTTTCGGAG 3810
 234 AspProGluArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGly 250
 |||||
 3811 GATCCCCAAGGCTCCAGGGGTTGGTGGAGTCCCTTATGTTCTCTCACCA 3860
 250 nProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThr 267
 |||||
 3861 GCCTACTTGGATGATGTCACACAGCTGCTGACAGCATCTTTCACACCG 3910
 267 LuGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAla 283
 |||||
 3911 AGAGGCGAGAGAAATCTATAGAGCTAGAAAAAATGTTCTCGGGGCC 3960
 284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300
 |||||
 3961 GACGGCGCACCCCGGCTTGCAAAATGATGATGACATGGGATTTCCCTT 4010
 300 uThrArgProGlyTyrAspTyrAsnThrAlaGlyAlaArgLysSerLeu 317
 |||||
 4011 AACTCGCCCCGGTGGACTACACACAGCGCTGAAGATGAGGAGAGCTTGA 4060
 317 ySLeTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
 |||||
 4061 AAATTTATCGCCAGGCTCTGTGGCGGGTCTCCGGGCGCTCAAGACGG 4110
 334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPro 350
 |||||
 4111 CCACATATTTGGCTAAGTAAAGAAAGTATGACAGGACGGAATGAACC 4160

```

350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrP
|||||
4161 CCCCCCTGTTTTCTTGAGAGGCTCTTGAGAGCTTCAGCGGTACACCC 4210
367 roPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe
|||||
4211 CTTTATCTCCACCTCAGAGGCCCAAAAAGCTCAGTGGCTTGCGCTTT 4260
384 lIleGlySerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu
|||||
4261 ATAGGACATCAGCCCTTGATATTAAGAAAGAGCTTCAGAGCTGGAAG 4310
400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysVal
|||||
4311 GTTACAGAGGCTGAGTACGTGATCTACTGAAGGAGCGAGAGAAAGTAT 4360
417 yTTrArgArgGluThrGluGluGluLysGluGlnArgLysGluLysGlu
|||||
4361 ATTACAAAAGGAGACAGAGAAGAAAGAAAGAAAGAAAGAGAGAGAA 4410
434 ArgGluGluArgGluGluArgArgAspArgArgGlnGluLysAsnLeuTh
|||||
4411 AGAGAGAAAGGAGAGAGAAAGACGTATTAACGCAAGAGAGAAATTTGAC 4460
450 rIysIleLeuAlaAlaValAlaGluGlyLysSerSerArgGluArgGlu
|||||
4461 TTAGATCTTGCGCTGCGAGTGTGAAGGAGAAAGCAATACGAAAGAGAA 4510
467 rGAspPheArgLysIleArgSerGlyProArgGlnSerGlyLysnLeuGly
|||||
4511 GAGATTTAGAAAATTAGTCAAGGCCCTCAGACAGTCAAGGAACTGGGC 4560
484 AsnArgThrProLeuAspLysAspClnCysAlaTyrCysLysGluLysGlu
|||||
4561 AATAGACCCCACTGACAGAGACCAATGTGATATTGTAAGAAGAGAG 4610
500 yHISThrAlaArgAsnCysProLysGlyLysnLysGlyProLysValL
|||||
4611 ACACGTGGCAGAGAACTGCCCAAGAGAGAAAGAGAGCAAGAGATCC 4660
517 eAlaLeuGluGluAspLysAsp 524
4661 TAGCTCTAGAGAGATTAAGAT 4683

```

seq_name: /cgn2_6/ptodata/2/lna/6a_COMB.seq:US-09-075-272-1

seq_documentation_block:

Sequence 1, Application US/09075272

Patent No. 6136598

GENERAL INFORMATION:

APPLICANT: MILLER, A. DUSTY

APPLICANT: WOLGAMOT, GREG

APPLICANT: BONHAM, LYNN

TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL

TITLE OF INVENTION: PACKAGING CELL LINES

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: U.S.A.

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/075,272

FILING DATE: 08-MAY-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 60/046,140
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: POOR, Brian W.
; REGISTRATION NUMBER: 32,928
; REFERENCE/DOCKET NUMBER: 14538A-003710
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8655 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-075-272-1

```

```

alignment_scores:
    Quality: 1684.00    Length: 541
    Ratio: 3.907        Gaps: 10
    Percent Similarity: 79.667    Percent Identity: 61.553

```

```

alignment_block:
US-09-171-553B-4 x US-09-075-272-1

```

```

Align seg 1/1 to: US-09-075-272-1 from: 1 to: 8655

```

```

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisThrPr 17
|||||
722 ATGGGACAACTGCTGCTCCACTCCCTTCTTAACCTCGAGCATGGAA 771
17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyLys 34
|||||
772 GGAGGTAAATTCAGAGCACACAAATCACTGAGTGAAGTGAAGAAAGCC 821
34 rOTPrGlnThrPheCysAlaSerGluThrProThrPheAspValGlyTrp 50
|||||
822 CATGGCAAACTTTTGGCCCTCCGAGTGGCCAAAGTTTGGAGTGGCTGG 871
51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaI 67
|||||
872 CCACCGGAGGGTGTCTTTGACTGTGACTGATCGCCCGCTGAGGCGAAT 921
67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
|||||
922 TGTTTTTCAGGAA...GAAGGGGTCAACCTGATCAATGCCCTTACATTG 968
84 eUTHrTrpGlnAspLeuAlaGluAspProProThrProValLysProTrp 100
|||||
968 TGACCTGGCAGAGTCTGCTCAGTCCAGTCCACCTGCTGCTGCTGCTG 1018
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyLys 117
|||||
1019 .....ACCCCAAAATCTTCGAATACTGAGGTTGAGTTGC 1053
117 sAsnLysHisSerAlaGluLysValGluProSerSer.....SerYrL 132
|||||
1054 CCACTCCGATGACAGTGAATTCGTGCTGCTGAGCAAGAGATTT 1103
132 eUPrOArgAspArgGlyAlaAlaAspLeuAlaGluThrProThrCysSer 148
|||||
1104 ATCCAGAGATTGAGACCTCTCTGATGATGATGATCCCAAGCTCCCTTAC 1153
149 ProThrProLeuSerSerThrGlyCysCysGluGlyThrSerAlaPro.. 164
|||||
1154 CCCCCTGCCAGCAGCCACCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1203
165 .....ProGlyAlaProValAlaGluGlyProAlaAlaGlyThrArg 179
|||||
1204 GCCGGAACCAACAGCAGAGCGGGGCTCAAGGAGAGAGAGAGAGAGAGAG 1253

```

```

179 eArGArgGLyAlaThrProGluArg.....ThrAspGluIleAla 192
|||||.....|
1254 GCCGCCGGGGTCGAGCCCTGACAGAGGAGGGGCCCAATTCCACACTT 1303
193 IleuProLeuArgThrTyr...GlyProPomMetProGly...GlyGI 207
|||||.....|
1304 GCCCTGGCCCTTAGAGCTCATGTGGAGGGCCCAACGCCAGACCTAATGA 1353
207 nLeuGlnProLeuGlnTyrTTPProPheSerSerAlaAspLeuTyrAsnT 224
|||||.....|
1354 TCTCATCTCTTACAGTACTGCTTTCTTCTCTGATTATATATAT 1403
224 rPLyThAsnHisProPheSerGluAspProGlnArgLeuThrGly 240
|||||.....|
1404 GGAAGAACACACCCCTCTCTGAGAAACCCCTGCGCTACTGCG 1453
241 LeuValGluSerIleuMetPheSerHisGlnProThrTrpAspAspGly 257
|||||.....|
1454 CTCCTTGAAATCCTTATGTTCTCTCATCAGCCCACTTGGATGATTCGA 1503
257 nGlnLeuLeuGlnThrLeuPheThrThrGluGluArgGluArgIleLeu 274
|||||.....|
1504 GCAGCTTTGACAGTTCTCTTACACAGAGGAAAGGAAAGAAATCCTGA 1553
274 eugGluAlaArgLysAsnValProGlyAlaAspGlyArgProThrGlnLeu 290
|||||.....|
1554 TGGAGGCAAGAAAAAGCTTCTAGAGAGGAGGACACCACTGCTCTC 1603
291 GlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTrpAspTyr 307
|||||.....|
1604 CCTAACCTCTGAGAGAGCTTCCCTTGAACCCCTCAACTGGGACTA 1653
307 rAsnThrAlaGluArgGluSerLeuLysIleTyrArgGlnAlaLeuVal 324
|||||.....|
1654 CAACACTGGGAGAGTAGGGAGCGCTCTCTTACCGTGAGACTGTGG 1703
324 aAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340
|||||.....|
1704 TGGCAGGCTCTCAGAGAGCCGCAAGACGCCCAATTTGGCTAAGGTA 1753
341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArg 357
|||||.....|
1754 AGAGAGGCTTCTCAGAGGAGCACTAACCCCTGATCTTCTTGAGAGC 1803
357 GlnuMetGluAlaPheArgArgPheThrProPheAspProThrSerGlu 374
|||||.....|
1804 TCTTATGAGGCGGTATAGAGGTACACCCCTTTCATCCCTGCTCAGAG 1853
374 IagGlnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAsp 390
|||||.....|
1854 GACAGAAAAGCCGCTGATGACCATGCTTATGGCCACTGCTCCGAT 1903
391 IleArgLysLysLeuGlnArgLeuGlnGluLeuGlnGluAlaLeuArg 407
|||||.....|
1904 ATTAAGAAAAGCTTACAGAGGCTGAGAGGCTCCCAAGATTTATCCCTCCA 1953
407 gAspLeuValArgGluAlaGluLysValTyrTyrArgArgGlnThrGlu 424
|||||.....|
1954 ACATTATAGTGAAGAGAGAGAGAGGTTTATCACAAGAGAGAAACAGAG 2003
424 IugLysGluGlnArgLysGluLysGluArgGluGluArgGluGluArg 440
|||||.....|
2004 AAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2053
441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaIleValVa 457
|||||.....|
2054 CGAGATCGCCGTCAGAGAGAGAAATTTGATGATTTGCGCGCTGAT 2103
457 IgLysGlyLysSerSerArgGluArgGluArgAspPheArgLysIleArgS 474
|||||.....|
2104 AAATGATGAGCAGCTCAGAAAAA..... 2125
474 eGlyProArgGlnSerGlyAsnLeuGlyAsnArg..... 485

```

```

2126 .....GGCAGACAGATTCCTGCGCAACAGGGGACAGAAACGCCCA 2167
486 .....ThrProLeuAspLysAspGlnCysAlaTyrCysLysGI 498
2168 GTGGCAGAAAGACCGCGCTGAGAAAAAGACCAATGCGCTTTGCAAAAG 2217
498 uLysGlyHisTTPAlaArgAsnGlyProLysGlyLysGlyProL 515
2218 AAAAGACACTGGGCTAAAGACTGCGCTTAAGAAA...AGAAGCAATTC 2264
515 yValLeuAlaLeuGluGluAsp 522
2265 AGGTCTGACCTAGAGAGAGAT 2287

```

```

seq_name: /cgn2_6/ptodata/2/lna/6A.COMB.seq:US-08-716-351A-1

```

```

seq_documentation_block:
; Sequence 1, Application US/08716351A
; Patent No. 6033905
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Glibon Ape Leukemia Virus-Based
; TITLE OF INVENTION: Retroviral Vectors
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,351A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03784
; FILING DATE: 06-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 15280-128-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; FAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..8535
; OTHER INFORMATION: /standard_name="Galv SEATO Genome"
US-08-716-351A-1

```

```

alignment_scores:
Quality: 1660.50 Length: 542
Ratio: 3.944 Gaps: 10
Percent Similarity: 77.675 Percent Identity: 61.439

```

```

alignment_block:
US-09-171-553B-4 x US-08-716-351A-1

```

```

Align seg 1/1 to: US-08-716-351A-1 from: 1 to: 8535

```

```

1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTyrPrh 17
|||||.....|
1001 ATGGGACACAGATATTTCTACCCCTATCTCCCTCAGCTTAATATCAGCGAG 1050

```



```
17  rgluValArgSerArgAlaHisAsnLeuSerValGluValLysLysGlyP 34
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1051 AGATGTGAGACAGAGGGCTCACATCTATCCGTGAAATCAAAAGGGA 1100
34  rortpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1101 AATGGAGACTTCTGTCTCCGAGTGGCCACATCCGCTGGGGTGG 1150
51  ProSerGluGlyThrPheAsnSerGluIleLeuAlaValLysAlaIle 67
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1151 CCACGGAGGAGAACTTTATCTCTCTCATCTTTTGGATTAAGAAAT 1200
67  ellePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1201 TGTCTTTCAGGAGAAC...GGGGGACATCCGAGCAAGTTCATATATCG 1247
84  eutThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1248 TGGTATGGAGAGACCTCGCCAGAAATCCGCCACCATGGGTG... 1288
101  LeuAsnLysProArgLysProGlyProArgLysLeuAlaLeuGlyGlu 117
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1289 .....CCAGCCTCCGGCAAGTGGCTGTGTCTGTGATAC 1323
117  sasnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1324 CCGAGACCAAGTTCGGGGAGG...CCATCAGCTCTCCCGGACCC 1367
134  rGAspArgGlyAlaAlaAsp.....LeuAlaGlyThrProThr... 146
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1368 CCATCTACCCGGCAACAGACGACTTACTCTCTGTGAACCCAGGCC 1417
147  .....CysSerProThrProLeuSerSerThrGlyCysCys 158
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1418 CCGCCCTATCCGGGGGCACTGGCACCCCTCTGGCCCTCAGGCATGG 1467
158  setGluThrSerAlaProProGlyAlaProValAlaGluGlyProAlaA 175
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1468 ACCGCCCTGCAGCCAGATGCCGAGATAGCATCTGAGGGCCAGCCG 1517
175  IAGLYThrArgSerArgArgGlyAlaThrPro.....GluArgThr 188
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1518 CTGGGACCAAGAGTCCGCGTCCGACGACGACAGACACTCGGTCCT 1567
189  AspGluIleAlaIleLeuProLeuArgThrTyrGlyProPromeProG 205
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1568 GACTCCAGCTGATTTGGCCCTCGAGCAGATAGSACCCCGCGGACCC 1617
205  yGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAspLeu 222
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1618 CAATGGCTGTCCCTCTACAAATATTTGCCCTTTTCTCAGCAGATCT 1667
222  yfAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeu 238
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1668 AATATTGGAAATCTATATCATCCCTTTTCTGAAAACCCAGCAGTCTC 1717
239  ThrGlyLeuValGluSerLeuMetPheSerHisGlnProThrTrpAsp 255
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1718 ACGGGGCTCCTTGAGTCTCTATATGTTCCCATCAGCCACTTGGAGCA 1767
255  pCysGlnGlnLeuLeuGlnThrLeuPheThrThrGluGluArgGluArg 272
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1768 TTGCCAAGAGCTCTACAGATCTTTTACCACTGAGGAGAGGGAAGAA 1817
272  IeLeuLeuGlnAlaArgLysAsnValProGlyAlaAspGlyArgProThr 288
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1818 TTCTCTGAGGCCCCGCAAAATGTCTTGGGAGCAATGGGGCCCTACAA 1867
289  GlnLeuGlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTr 305
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1868 CAGCTCGAAGAACCTTAAATAGAGGCTTCCCGCCATCATCAGCTCACTG 1917
305  pAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArgGlnA 322

1918 GGAATTACAAACAGCCGAGTAGGAGACCTCTTGTGATACGCCGGA 1967
322  IAlaLeuValAlaGlyLeuArgGlyAlaSerAlaGatGProThrAsnLeuAla 338
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1968 CTTAGTGGCAGGTCTCAAAAGGGAGACTCGCTCTCTACCAATTTGGCT 2017
339  LysValArgGluValMetGlnGlyProAsnGluProProSerValPheLe 355
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2018 AAGGTAAAGAGGCTCTTGACAGGAGCCGCGAAGACCCCTTGCTTCTT 2067
355  uGluArgLeuMetGluAlaPheArgArgPheThrProPheAspProThr 372
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2068 AGAAGCCCTGATGAGAGGCTTATAGAGATACACTCGCTTGATCTCTT 2117
372  erGluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAla 388
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2118 CTGAGGGACAACAGGCTGCGGTCCCATGGCTTATTCGACATCGAGCC 2167
389  LeuAspIleArgLysLysLeuGlnArgLeuGlnGlyLeuGlnGluAlaG 405
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2168 CCAGATATTCAGAAATACTTACAGAGCTAGAGGGCTCCAGGACTTTC 2217
405  uLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArgGlu 422
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2218 CTTCAGAGATTTACTAAAGAGGCAAAAGGTCTACCATTAAGAGAGA 2267
422  hrGluGluGluLysGluGlnArgLysGluLysGluValGluGluGlu 438
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2268 CAGAAAGAAAGAAAGCAAGAAAGAAAGAAAGGAGGAGAAAGAAAGAG 2317
439  GluArgArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaI 455
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2318 AGGGGGGCGCATTAGCCGAGCAAGAAAAAAGTTGACTTAAATTTGCGCCG 2367
455  aValValGluGlyLysSerSerArgGluArgGluArgAspPheArgLys 472
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2368 AGTAGTA.....AGTAGAGAA..... 2383
472  IeArgSerGlyProArgGlnSerGlyAsnLeuGlyAsn..... 484
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2384 ..GGTCCACAGCTAGGCAAGACAGCAAGCACTGAGCAACGAGCAAGAG 2431
485  .....ArgThrProLeuAspLysAspGlnCysAlaTyrCys 496
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2432 ACACCTTAGGAGTGAAGAACTCCACTAGACAAAGACCAAGTGGCATCTG 2481
496  sLysGluLysGlyHisTrpAlaArgAsnCysProLysLysGlyAsn...L 512
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2482 TAAAGAGAAAGGCGCATTTGGCAAGAGAAATGCCCGGAAAAAAGACAGCTCA 2531
512  ySGlyProLysValLeuAlaLeuGlu 520
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2532 GAGAAAGCAAGGTTCTAGCCCTAGAT 2557

seq_name: /cgn2_6/plodata/2/ina/6B_COMB.seq:US-09-011-745-3
seq_documentation_block:
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiko
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
```

```
? NUMBER OF SEQ ID NOS: 29
? SOFTWARE: Patent In Ver. 2.0
? SEQ ID NO 3
? LENGTH: 7308
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: Portion of
? US-09-011-745-3

alignment scores:
  Quality: 1355.50      Length: 578
  Ratio: 3.440          Gaps: 15
  Percent similarity: 68.166      Percent Identity: 48.962

alignment block:
US-09-171-553B-4 x US-09-011-745-3 ..

Align seg 1/1 to: US-09-011-745-3 from: 1 to: 7308

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTyrPth 17
  |||||
1567 ATGGCCAGACTGTTACCACTCCCTTAAGTTGAGCTTAGGCTACGTGAA 1616
  |||||
17 GlnValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
  |||||
1617 AGATGTCGAGCGGATCGCTCACACCAGCTCGGTAGATGTCAAGAGAAC 1666
  |||||
34 roTgriGlnThrPheCysAlaSerGluTyrProThrPheAspValGlyTyr 50
  |||||
1667 GTTGGGTAACTTGTCTGTGCAAGATGCGCAACCTTTAACGTGAGATGG 1716
  |||||
51 ProSerGlnGlyThrPheAsnSerGluLeuLeuAlaValLysAlaIle 67
  |||||
1717 CCGCAGAGCGACCTTTAACCGAGACCTCATCACCCAGGTTAAGATCAA 1766
  |||||
67 eilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
  |||||
1767 GGTCTTTTACCTGCGCCGATGACACCCAGACCGAGTCCCTCATGTCG 1816
  |||||
84 euThrTyrGlnAspLeuAlaGluAspProProTyrValLysProTyr 100
  |||||
1817 TGACCTGGGAGCCTTGGCTTTTGACCCCTCCCTGGGTCAAGCCCTTT 1866
  |||||
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyLys 117
  |||||
1867 GTACAC...CTTAAGCTTCGGCTCT... 1890
  |||||
117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuPro 134
  |||||
1891 .....CTTCTCCATCCGCGCCGCTCTCCCTCC 1918
  |||||
134 rGAsp.....ArgGlyAlaAlaAspLeuAlaGly 143
  |||||
1919 TTGAACCTCTCTGTCGACCCGCGCTCGATCTCTCTTTATCCACCCCTC 1968
  |||||
144 ThrProThrCysSerProThrPro.....LeuSerSerThrG 156
  |||||
1969 ACTCTCTCTAGGCGCCAAACCTTAACCTCAAGTTCTTTCGACAGTGG 2018
  |||||
156 yCysCysGlnGly.....ThrSerAlaProPro... 165
  |||||
2019 G.....GGGCGGCTCATGACCTACTATACAGAGACCCCGCGCTT 2059
  |||||
165 ..... 165
2060 ATAGGACCCCAAGACCCCTTCGACAGGAGGAGAAATGGTGAAGAA 2109
  |||||
166 GlnAlaProValValGlnGlyPro.....AlaAlaGlyTh 177
  |||||
2110 GCGACCCCTTGGGAGAGGCGACCGAGACCCCTCCCAATGGCATTCGCT 2159
```

```
177 rArgSerArgArgGlyAlaThrProGluArgThrAspGluIleAlaIle 194
  |||||
2160 ACGTGGGAGAGCGGAGCCCTGTGTGCGGACTCTCCACTACCTGCGACGCA 2209
  |||||
194 euProLeuArgThrTyrGlyProProMetProGlyGlnLeuGlnPro 210
  |||||
2210 TCCCCCTCCCGCAGAGAGA.....AACGACAG..... 2238
  |||||
211 LeuGlnIleThrProPheSerSerAlaAspLeuTyrAsnTyrLysThrs 227
  |||||
2239 CTTCATACTGCGCGCTTCTCTCTCTCTCTGACTTACACCTGGAATAAATA 2288
  |||||
227 nHisProProPheSerGluAspProGlnArgLeuThrGlyLeuValGln 244
  |||||
2289 TAACCTCTCTTTTCTGAAGATCCAGGTAAACGACGACGCTGATGAGAT 2338
  |||||
244 euLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
  |||||
2339 CTGTTCATCACCCATCACCCCACTCGGACGACGCTGACAGCTGTG 2388
  |||||
261 GlnThrLeuPheThrThrGluArgGluArgIleLeuLeuGluAlaAr 277
  |||||
2389 GGGACTCTCTGACCGGAGAAACAAACGSGTCTCTTAGAGGCTAG 2438
  |||||
277 gLysAsnValProGlyAlaAspGluArgProThrGlnLeuGlnAsnGlu 294
  |||||
2439 AAAGCGGCTCGGGCGCATATGCGGCCCTCACTCAAGCTGCCAATGAAG 2488
  |||||
294 leaSPMeGlyPheProLeuThrArgProGlyTyrAspTyrAsnThrAla 310
  |||||
2489 TCGATGCGCGCTTTCCTCCCTGAGCGCCACAGCGGATTAACACCCAG 2538
  |||||
311 GlnGlyArgGlnSerLeuLysIleTyrArgGlnAlaLeuValAlaGly 327
  |||||
2539 GCAGGTAGAAACCACTAGTCCACTATCCGCACTTCTCTTACCGGCTCT 2588
  |||||
327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluVal 344
  |||||
2589 CCAAAACGCGGGGAGAGCCCAACCAATTTGGCCAAAGTTAAAGAAATA 2638
  |||||
344 eTGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
  |||||
2639 CACAAGGCGCCATGATCTCCCTCGCGCTCTCTAGAGAGACTTAAGGAA 2688
  |||||
361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
  |||||
2689 GCGTATCGAGGTAACTCTTATGACCCCTGAGGACCCAGGCGCAAGAAAC 2738
  |||||
377 aSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLys 394
  |||||
2739 TAAATGTCTATGCTCTTCAATTGGCAGTCTGCGCCACAGACATTGGAGAA 2788
  |||||
394 yLsLeuGlnArgLeuGlnGlyLeuGlnGluAlaLeuAspLeuVal 410
  |||||
2789 AGTTAGAGAGGTTAGAGATTTAACAAACAGCGCTTGAGATTTGGT 2838
  |||||
411 ArgGlnAlaGluLysValTyrTyrArgArgGluThrGluGlnGluLysG 427
  |||||
2839 AGAGAGCAGCAAAAGATCTTAATTAACGAGAAACCCCGGAGAAAGAGA 2888
  |||||
427 uGlnArgLysGluLysGluArgGluGluArgGluArgGluArg..... 440
  |||||
2889 GGAACGTATCAGAGAGAAACAGAGGAGAAAGAAAGCGCGTAGAGACAG 2938
  |||||
441 .....ArgAspArgArgGlnGlnLysAsnLeu 449
  |||||
2939 AGGATGAGCAGAAAGAGAAAGAAAGATCTGAGAGACATAGAGAGATG 2988
  |||||
450 ThrLysIleLeuAlaAlaValValGlnGlyLysSerSerArgGlnArg 466
  |||||
2989 AGCAAGCATTTGGCCACTGTGCTTAGTGAAGAGAGAAACAG..... 3027
```

```

466 uatgasphearglyleargserglyproarglnserglyasnleug 483
      |||||.....
3028 .....CATGACAGGAGAGACAA...C 3046
483 lyasnargthrproleuasplasnalsalatriscylslysgly 499
      |||||.....
3047 GAAGGAGGTCCCAACTGATCGGACAGGTGCTACTGCAAGCAAAAG 3096
500 glyhistrpalargasnrcysprolysgly.....glyasnlysglypr 514
3097 GGGCACTGGGCTAAAGATTGTCACAAAGAACACAGAGACTCGGGGACC 3146
514 olys.....Valleualaleugluin 521
      |||||.....
3147 AAGACCCACGACCTCCCTCTGACGACCTAGATGAC 3180

seq_name: /cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-011-745-4

seq_documentation_block:
  Sequence 4, Application US/09011745
  Patent No. 6165715
  GENERAL INFORMATION:
  APPLICANT: Collins, Mary KL
  APPLICANT: Weiss, Robin A
  APPLICANT: Takeuchi, Yasuhiro
  APPLICANT: Cosset, Francois-Joic
  TITLE OF INVENTION: Expression systems
  FILE REFERENCE: 09/011,745
  CURRENT APPLICATION NUMBER: US/09/011,745
  EARLIER FILING DATE: 1998-06-22
  EARLIER APPLICATION NUMBER: PCT/GB96/02061
  EARLIER FILING DATE: 1996-08-23
  EARLIER APPLICATION NUMBER: GB9517263.1
  EARLIER FILING DATE: 1995-08-23
  NUMBER OF SEQ ID NOS: 29
  SOFTWARE: PatentIn Ver. 2.0
  SEQ ID NO 4
  LENGTH: 7308
  TYPE: DNA
  ORGANISM: Artificial Sequence
  FEATURE:
  OTHER INFORMATION: Description of Artificial Sequence: Portion of
  US-09-011-745-4

alignment_scores:
  Quality: 1355.50      Length: 578
  Ratio: 3.440          Gaps: 15
  Percent Similarity: 68.166      Percent Identity: 48.962

alignment_block:
  US-09-171-553B-4 x US-09-011-745-4 ..
  Align seg 1/1 to: US-09-011-745-4 from: 1 to: 7308

1 MetGlnGlnThrValThrThrProLeuSerLeuThrLeuAspHisThrPrh 17
|||||.....
1567 ATGGCCGAGCTGTACCACTCCCTTAAGTTGACCTTAGCTACACTGCGAA 1616
17 rglnvalargserargalaHisasnleuSerValGlnValLyslysglypr 34
|||||.....
1617 AGATTCGACGGGATCGCTCAACAACAGTCGATGATGTCACAGAGAC 1666
34 rotpglnthrPhcysalaserGlnThrProThrPhcaspvalGlyTyr 50
|||||.....
1667 GTTGGGTAACTTCCTGCTGAGAAAGGCCAACCTTTAACCTGCGATGG 1716
51 ProserGlnGlyThrPhcaspserGlnThrLeuAlaValLysalaail 67
|||||.....
1717 CCGCAGACGACCACTTTAACCGAAGCACTCAACCCAGGTTAAGATCAA 1766
67 ellepheglnthrGlyProGlySerHisProAspGlnGlnThrProTyrIlel 84

```

```

1767 GGTCTTTTACCTGGCCGATGACACCCAGACAGGCTCCCTACATCG 1816
      |||||.....
84 eutThrTrpGlnAspLeuAlaGlnAspProProThrProThrValLysProTrp 100
      |||||.....
1817 TGACCTTGGAAGCCCTTGCTTTTGACCCCTCCCTGAGGCTCAAGCCCTTT 1866
101 leuAsnLysProArgLysProGlyProArgIleleuAlaLeuGlyGly 117
      |||||.....
1867 GTACAC...CCTAAGCTCCGCTCCT..... 1890
117 sasnLysHisserAlaGlnLysValGlnProSerSerTyrLeuProA 134
      |||||.....
1891 .....CTTCCTCATCCGCGCCGCTCTCCCTCC 1918
134 rGAsp.....ArgGlyAlaAlaAspLeuAlaGly 143
      |||||.....
1919 TTGAAACCTCTCTGTCGACCCCGCTCGATCTCCCTTTTTCAGGCTTC 1968
144 ThrProThrCysSerProThrPro.....LeuSerSerThrGln 156
      |||||.....
1969 ACTCCTTCTTAGGGCCCAACCTAAACCTAAAGTTCTTTGTGACAGTGG 2018
156 yCysCysGly.....ThrSerAlaProPro.... 165
2019 G.....GGCCGCTCATGACACTACTTACAGAAAGACCCCGCTT 2059
165 ..... 165
2060 ATAGGAGCCCAAGACACCCCTTCGACAGGACGGAAATGTGGAGAA 2109
166 GlyAlaProValValGlyGlyPro.....AlaAlaGlyTh 177
      |||||.....
2110 GCGACCCCTGCGGAGAGGACCGACCCCTCCCAATGATCGATCGCT 2159
177 rArgSerArgArgGlyAlaThrProGluArgThrAspGlnIleAlaIlel 194
      |||||.....
2160 ACCTGGGAGACGGGAGCCCGCTGTGGCGACTACTCTCGCAGGACT 2209
194 eupProLeuArgThrTyrGlyProProMetProGlyGlnLeuGlnPro 210
      |||||.....
2210 TCCCCCTCGCGGAGAGGA.....AACGGACG..... 2238
211 leuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnTrpLysThrAs 227
      |||||.....
2239 CTTCATATACGGCGCTTCTCTCTCTGACTTACAACTGAAATAA 2288
227 nHisProProPheSerGlnAspProGlnArgLeuThrGlyLeuValGln 244
      |||||.....
2289 TAACCTTCTTTTTCGAAGATCCAGGTAAACGTGACAGCTGTGATGACT 2338
244 erLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260
      |||||.....
2339 CTGTTCATCACCACCATGACCCACCTGGGACGACACTGTCAAGCCTGTG 2388
261 GlnThrLeuPheThrThrGlnGlnArgGlnArgIleleuLeuGlnAlaAr 277
      |||||.....
2389 GGGACTGTGCTGACCGGAGAAAGAAACAGGGGTCTCTTAAGAGCTAG 2438
277 gLysAsnValProGlyAlaAspGlyArgProThrGlnGlnLeuAsnGln 294
      |||||.....
2439 AAGGCGGTGCGGGCGATGATGAGGCGCCCACTCACTCCCAATGAG 2488
294 leaSPMeGlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAla 310
      |||||.....
2489 TCGATGCGGCTTTTCCCTCGAGGCGCCAGACTGATGATACACACACAG 2538
311 GlnGlyArgGlnSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLe 327
      |||||.....
2539 CGAGACGCAACCACTAGTCCACTATGCGCAGTTCCTCCAGGGGCTCT 2588
327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGlnVal 344
      |||||.....

```

Quality:	1355.50	Length:	578
Ratio:	3.440	Gaps:	15
Percent Similarity:	68.166	Percent Identity:	48.962

Align seg 1/1 to: US-09-011-745-2 from: 1 to: 7616

CO: 05-09-011-145-2 FROM: 1 TO: 7616

1 MetGylGlnThrValThrThrProLeuSerLeuThrLeuAspHisIstrpH 17
 |||||
 28 ATGGCCAGACTGTITACACACTCCCTTAGTTGACCTTAGTACACTGGAA 77

```

17  rgUValArgSerAAlaHisnleuSerValGlnValLysLysGlyp 34
    :::::  :::::  :::::  :::::  :::::  :::::  :::::
78  AGATGTCAGCGGATCGCTCAACCAAGTCGCTAGATGTCAAGAAGAGAC 82

```

Proserfina) epithelial cells

34 roirrgnlnthrrheacysalaserlutrprothrphesapvalgltyr 50
||| ||||||| : : : : :
28 gttggggttacctctgcctccgaagtggcacaactttaacgtcgatcg 87

7 eilepgeglnthrglvpbrogivserhispacacactatg
 8 cgcgcagacgcgcacctttacccgaacaccttcacccagcttagatcna 927

4 euThrTTP6InaspleuInaGInaSPProProProTnrvInaIksProCTG 100
8 GGCTCTTTTACCTGCGCCGCATGCACACCCAGACCCAGGCTCCCTTACATCG 977

1 LeuAsnIlysProArgIyLysProGlyIProArgIleLeuAlaLeuGlyGlyI 100
8 TGACCTGGGGAAGCCCTTGGCTTTTGTACCCCCCTCCCTGGGTCAAGCCCTTT 102

7 sasnlYshlsserlaaglulYsaIgluProserSerTyrleuProA 134
8 GTACAC... CCTAAGCCTCCGCCCTCT 105

```

2      . . . . . CTTCTCTCAATCCGCCCCGTCTCTCCCCC 107
4  rgasp. . . . . ArgGlyAlaIaaAspLeuAlaGly 143
      . . . . .

```

TTGACCTCCTCTTGCACCCCGCCCTGCATCTCCCTTTATACACCCCTC	1129
ThiProThrcysSerProThrPro.....LeuSerSerThrG1	156

ACGTCCTCTCTTCAAGCGCCAAACCTAACCCTCAGTCTTTCTGCACATGG 1179
yCysCysglugly.....ThSerAlaProPro.... 165
- - - - - |||||
0888888888

[illegible][illegible]

```

TArgSerArgArgGlyAlaThrProGluArgThrAspLeuIleAlaIleIle 194
||||:|||||
ACGTGGAGACCGGAGCCCCCTGTGGCCGACGACCTACATCCCTGGAGGCAT 1370
:::
ACGTGGAGACCGGAGCCCCCTGTGGCCGACGACCTACATCCCTGGAGGCAT 1320

```

```

194 eupProLeuArgThrTyrglyProProMetProglyGlnLeuGlnPro 210
   :::::::::::::: ::::::::::::::
1371 TCCGCCCTCCCGCAGAGAGA.....AACGACAG..... 1399
211 LeuGlnTyrglyProProMetSerAlaAspLeuTyrglySerThrAs 227
   :::::::::::::: ::::::::::::::
1400 CTTCATATACGCGCTTCCTCTTCGACCTTACCAACGAGAAATAA 1449
227 nHisProProMetSerGlnAspProGlnArgLeuThrGlyLeuValGln 244
   :::::::::::::: ::::::::::::::
1450 TAACCTCTCTTTCTGAGAGTCCAGGTAACCTACAGCCTGATGTCAGT 1499
244 erLeuMetPheSerHisGlnProThrTrpAspPcySglnGlnLeu 260
   :::::::::::::: ::::::::::::::
1500 CTGTTTCATACACCATCAGCCACCTGGCAGCAGCTGTGACAGCTGTG 1549
261 GlnThrLeuPheThrThrGlnGlnArgGlnArgLeuLeuGlnValAar 277
   :::::::::::::: ::::::::::::::
1550 GGGACTCTGCTGACCGGAGAGAAACAAACAGGCTGCTTAGAGGCTAG 1599
277 gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGln 294
   :::::::::::::: ::::::::::::::
1600 AAAGCGGTGGCGGCGGCGATGATGGCGGCCCACTCACTGCCAATGAG 1649
294 leaSPmetGlyPheProLeuThrArgProGlyTrpAspTrpAsnThrAla 310
   :::::::::::::: ::::::::::::::
1650 TCGATGCCGCTTTCCTCGAGCGCCAGACTGGGATTACACACCCAG 1699
311 GlnGlyArgGlnSerLeuLysIleThrArgGlnAlaLeuValAlaGlyLe 327
   :::::::::::::: ::::::::::::::
1700 GCAGGTAGAGAACCACTAGTCCACTATGCGCGTTCCTCTCCAGCGGCT 1749
327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGlnValM 344
   :::::::::::::: ::::::::::::::
1750 CCAAAACGGGGGAGAGAGCCCAATTTGGCCAAAGGTAAGAAATAA 1799
344 etGlnGlyProAsnGlnProProSerValPheLeuGlnArgLeuMetIu 360
   :::::::::::::: ::::::::::::::
1800 CACAGGGGCCAATGAGTCTCCCTGCGCTTCCTGAGAGACTTAAGGAA 1849
361 AlaPheArgArgPheThrProPheAspProThrSerGlnAlaGlnLysAl 377
   :::::::::::::: ::::::::::::::
1850 GCCTATCGAGAGTACACTCTTATGACCTCGTAGAGCCAGGAGCAAGAAC 1899
377 aserValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysL 394
   :::::::::::::: ::::::::::::::
1900 TAATGTCTATGCTTCTTATTTGGCAGTCCGCCAGCACTTGGGAGAGA 1949
394 ySLeuGlnArgLeuGlnGlyLeuGlnGlnAlaGlnLeuArgAspLeuVal 410
   :::::::::::::: ::::::::::::::
1950 AGTTAGAGAGGTTAGAGATTAAACAAACAGCGTTGAGATTGGTT 1999
411 ArgGlnAlaGlnLysValTyrglyTrpArgArgGlnThrGlnGlnLysGln 427
   :::::::::::::: ::::::::::::::
2000 AGAGAGGACGAAAGATCTTATTAACGAGAAACCCCGAGAGAAAGAGA 2049
427 uGlnArgLysGlnLysGlnArgGlnArgGlnArg..... 440
   :::::::::::::: ::::::::::::::
2050 GGAAGCTATCTCAGAGAGAAACAGAGAAAGAGAGAGCGCTGAGACAG 2099
441 .....ArgAspArgArgGlnGlnLysAsnLeu 449
   :::::::::::::: ::::::::::::::
2100 AGGATGAGCGAAGAAAGAAAGAGATGCTGAGAGACATAGAGAGATG 2149
450 ThrLysIleLeuAlaAlaValAlaGlnGlyLysSerSerArgGlnArgGln 466
   :::::::::::::: ::::::::::::::
2150 AGCAAGCTATTTGGCAGCTGCTGAGGAGAGAGAAACAG..... 2188
466 uArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuG 483
   :::::::::::::: ::::::::::::::
2189 .....GATAGCAAGGAGAGAA...C 2207
483 LysAsnArgThrProLeuAspLysAspGlnCysAlaTyrglyCysLysGlnLys 499

```

```

2208 GAAGGAGGTCCCACTCGATCGGAGACGAGTGTGCTACTGCAAGAGAAAG 2257
   :::::::::::::: ::::::::::::::
500 GlnHisTrpAlaArgAsnGlyProLysLys.....GlyAsnLysGlyPr 514
   :::::::::::::: ::::::::::::::
2258 GGGCACTGGGCTAAAGATTGTCCCAAGAAACAGAGAGACCTCGGGGACC 2307
514 oLys.....ValLeuAlaLeuGlnGln 521
   :::::::::::::: ::::::::::::::
2308 AAGACCCGAGACCTCCCTCTGACCTAGATGAC 2341
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-08-850-961-1
seq_documentation_block:
/ Sequence 1, Application US/08850961
/ Patent No. 6013517
/ GENERAL INFORMATION:
/ APPLICANT: Respass, James G.
/ APPLICANT: De Polo, Nicholas J.
/ APPLICANT: Chada, Sunil
/ APPLICANT: Sauter, Sybille
/ APPLICANT: Bodner, Mordechai
/ APPLICANT: Driver, David A.
/ TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
/ NUMBER OF SEQUENCES: 45
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Chiron Corporation, Intellectual Property - R440
/ STREET: P.O. Box 8097
/ CITY: Emeryville
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94662-8097
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/850,961
/ FILING DATE: 05-MAY-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kruse, No. 6013517/man J.
/ REGISTRATION NUMBER: 35,235
/ REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (510) 601-3520
/ TELEFAX: (510) 655-3542
/ INFORMATION FOR SEQ. ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 8332 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-08-850-961-1
alignment_scores:
Quality: 1355.50 Length: 578
Ratio: 3.440 Gaps: 15
Percent Similarity: 68.166 Percent Identity: 48.962
alignment_block:
US-09-171-553b-4 x US-08-850-961-1 ..
Align seg 1/1 to: US-08-850-961-1 from: 1 to: 8332
1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
   :::::::::::::: ::::::::::::::
621 ATGGGCGACACTGTATACCACTCCCTTAAGTTTGAACCTTAGGCACTGGAA 670
17 rGlnValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGly 34
   :::::::::::::: ::::::::::::::

```

```

671 AGATGTCGAGCGATCGCTCACACACCGTGTAGATGTCAAGAGAC 720
34  rofThrGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
721 GTTGGGTACCTTCTGCTCTGAGATGCGCAACCTTTAAGTCGATGG 770
51  ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIle 67
771 CCGGAGACGCGACCTTTAACCGAGACCTCATCACAGGTAAAGATCAA 820
67  eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleIle 84
821 GGTCTTTTACCTGGCCCGCATGAGACACCGACAGAGTCCCTACATCG 870
84  eutThrTrpGlnAspLeuAlaGluAspProProPOTrPValLysProTrp 100
871 TGACCTGGGAAGCCTTGCTTTTGAACCCCTCCCTGCTCAAGCCCTTT 920
101  LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGly 117
921 GTACAC...CCTAAGCTCCGCGCTCT... 944
117  sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
945 .....CTTCCTCATCCGCGCGCTCTCTCC 972
134  rGAsp.....ArgGlyAlaAlaAspLeuAlaGly 143
973 TTGAACCTCCGCTGTCAGCCCGCTGAGCTCTTATTCACGCGCTC 1022
144  ThrProThrCysSerProThrPro.....LeuSerSerThrS1 156
1023 ACTCCCTCTCTAGCGCCCAACCTAAACCTCAAGTCTTTCGACAGTGG 1072
156  yCysCysGluGly.....ThrSerAlaProPro... 165
1073 G.....GGCCCGCTATGACCTACTACTACAGAAACCCCGCTT 1113
165 ..... 165
1114 ATAGGACCCCAAGACACCCCTTCGACAGGAGGAAATGGTGAGAA 1163
166  GlyAlaProValValGluGlyPro.....AlaAlaGlyTh 177
1164 GCGACCCCTGCGGAGAGGACCGACCCCTCCCAATGGCATCTCGCT 1213
177  rArgSerArgArgGlyAlaThrProGluArgThrAspGluIleAlaIle 194
1214 AGCTGGGAGACGGAGACCCCTGTGGCGACTCCACTCTCGCAGGCA 1263
194  eutProLeuArgThrTyrGlyProPrometProGlyGlnLeuGlnPro 210
1264 TCCCTCCGCGCAGAGAA.....AACGGACAG..... 1292
211  LeuGlnIyrTrpProPheSerSerAlaAspLeuTyrAsnTrpLysThrS 227
1293 CTTCAATAGTGGCGCTTCTCTCTGACCTTTAACAATGGAAATAA 1342
227  nhIsProProPheSerGluAspProGlnArgLeuThrGlyLeuValGlu 244
1343 TAACCTCTTTTCTGAGATCCAGGTAAGTACAGACTGTGATCGAG 1392
244  erLeuMetPheSerHisGlnProThrTrpAspCysGlnGlnLeuLeu 260
1393 CTGTCTCATCACCATCAGCCACCTGGGAGCAGCTCAGCAGCTGTT 1442
261  GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAla 277
1443 GGGACTCTGCTGACCGGAGAGAAACAGGGTCTCTTAAAGAGCTAG 1492
277  gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGln 294
1493 AAAGGCGGTGGGGGCGATATGGGCGCCCACTCAACTGCGCAATGA 1542

```

```

294  LeAspMetGlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAla 310
1543 TCGATGCCGCTTCCCTCCAGGCGCCAGCTGGATTCACACCCAG 1592
311  GluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLe 327
1593 GCAGGTAGAACCACTAGTGCCTATTCGCCAGTTCCTCTACGCGCT 1642
327  uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluVal 344
1643 CCAAAACCGCGGAGAGCCCAACCAATTGGCCAAAGTAAAGAAATA 1692
344  eGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
1693 CACAAAGGCGCAATGACTCTCTCGCGCTTCTTACAGAACCTTAAG 1742
361  AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
1743 GCTATCGCAGGTACACTCTTATGACCTGAGAGACCCAGGCGAAGAA 1792
377  aSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLys 394
1793 TAATGTTCTATGCTTTCATTGGCAGCTCGCCAGCATTTGGAGAA 1842
394  yLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuVal 410
1843 AGTAGAGAGAGCTTAGAAGATTAAAAACAGACGCTTGAGATTTGGT 1892
411  ArgGluAlaGluLysValTyrTyrArgArgGluThrGluGluGlyS1 427
1893 AGAGAGCAAAAGATCTTAAATAACGAGAAACCCCGAAGAAAGAA 1942
427  uGlnArgLysGluLysGluArgGluGluArgGluGluArg... 440
1943 GGAAGCTATCAGGAGAGAAACAGAGAAAGAAAGACCGCTAGACAG 1992
441 .....ArgAspArgArgGlnGluLysAsnLeu 449
1993 AGCATGACGAGAAAGAGAAAGAGATCGTAGAGACATRAGAGATG 2042
450  ThrLysIleLeuAlaValValGluGlyLysSerSerArgGluArg1 466
2043 AGCAACCTATTGGCCACTGCTGTAGTACAGAAACAG..... 2081
466  uArgAspPheArgLysIleArgSerGlyProArgGlnSerLysLeuG 483
2082 .....GATAGACAGGAGAGAGAA...C 2100
483  LysAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLys 499
2101 GAAGAGAGTCCCACTGATCGCGCAGCTGTGCTCTACTCTCAAGAAAG 2150
500  GlyHisThrPalaArgAsnCysProLysLys.....GlyAsnLysGlyPr 514
2151 GGGCACTGGCTTAAAGATTGCCAAGAAACACAGAGACCTGGGGAGC 2200
514  Olys.....ValLeuAlaLeuGluGlu 521
2201 AAGACCCGAGACCTCCCTCGACCTTAGATGAC 2234

```

seq_name: /cgn2_6/prodata/2/lna/5A.COMB.seq:US-08-258-420-13

seq_documentation_block:

Sequence 13, Application US/08258420

Patent No. 5710037

GENERAL INFORMATION:

APPLICANT: Menhais, Arthur W.

APPLICANT: Vanin, Elio F.

TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Ve

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carella, Byrne, Bain, Gillillian, Cecchi, Stewart & Olstein

```

? STREET: 6 Becker Farm Road
? CITY: Roseland
? STATE: New Jersey
? COUNTRY: USA
? ZIP: 07068
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 inch diskette
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WordPerfect 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/258,420
? FILING DATE: 10-JUN-1994
? CLASSIFICATION: 435
? PRIORITY APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Olstein, Elliot M.
? REGISTRATION NUMBER: 24,025
? REFERENCE/DOCKET NUMBER: 271010-208
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-994-1700
? TELEFAX: 201-994-1744
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8202 bases
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: genomic DNA
? FEATURE:
? NAME/KEY: viral genome
? US-08-258-420-13

```

```

alignment_scores:
  Quality: 1354.50      Length: 578
  Ratio: 3.438          Gaps: 15
  Percent Similarity: 68.166  Percent Identity: 48.789

```

alignment_block:

US-09-171-553B-4 x US-08-258-420-13/rev ..

Align seg 1/1 to reverse of: US-08-258-420-13 from: 1 to: 8202

```

1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrp 17
  |||||
7146 ATGGCCGAGCTGTTACCACTCCCTTAAGTTGACTTACCTGAA 7097
  |||||
17 rGlValArgSerArgAlaHisAsnLeuSerValGlnValLysGly 34
  |||||
7096 AGATGTCAGCGGATCGCTCACACACAGTCGATGATGTCAGAGAGAC 7047
  |||||
34 rortGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
  |||||
7046 GTTGGGTACTCTCTGCTGCATGACAGTGGCAACCTTAACATGATGG 6997
  |||||
51 ProSerGluGlyThrPheAsnSerGluLeuLeuValLysAlaIle 67
  |||||
6996 CCGGAGAGCGCACCTTTAAACCGAGACCTCATCCAGGTTAAGATCAA 6947
  |||||
67 eilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
  |||||
6946 GGTCTTTTACCTGGCCCGCATGGAGACCCAGACAGTCCCTACATCG 6897
  |||||
84 eurtTrpGlnAspLeuAlaGluAspProProTyrValLysProTyr 100
  |||||
6896 TGACCTGGAGAGCTTGGCTTTTACCCCTCCCTGGGTCAAGCCCTTT 6847
  |||||
101 LeuAsnLysProArgLysProGlyProArgLysLeuAlaLeuGlyGly 117
  |||||
6846 GTACAC...CCTAAGCTTCGCTCTCT..... 6823

```

```

117 sAsnLysHisSerAlaGluLysValGluProSerSerSeryrLeuPro 134
  |||||
6822 .....CTTCCTCCATCCGCGCCGCTCTCC 6795
  |||||
134 rGAsp.....ArgGlyAlaAlaAspLeuAlaGly 143
  |||||
6794 TTGAACCTCTCGTTGACCCCGCCGCTCTTATCCAGCCCTC 6745
  |||||
144 ThrProThrCysSerProThrPro.....LeuSerSerThrG1 156
  |||||
6744 ACCTCTCTCTAGGCGCAACCTTAACCTCAAGTTCTTTCGACAGTGG 6695
  |||||
156 yCysGlyGly.....ThSerAlaProPro... 165
  |||||
6694 G.....GGGCGCTCATGACCTACTTACAGAGAGCCCGGCTT 6654
  |||||
165 ..... 165
6653 ATAGGAGCCCAAGACCCCTCCGACAGGAGGAGAAATGATGAGAA 6604
  |||||
166 GlyAlaProValValGluGlyPro.....AlaAlaGlyThr 177
  |||||
6603 GCGACCCCTCGGAGAGAGACCGACCCCTCCCAATGGCATCTCGCT 6554
  |||||
177 rArgSerArgArgGlyAlaThrProGluArgThrAspGluAlaIle 194
  |||||
6553 ACCTGGAGAGAGGAGACCCCTGTGGCGAGTCACACTCTCGAGCAT 6504
  |||||
194 eudProLeuArgThrTrpGlyProPrometProGlyGlyGlnLeuGln 210
  |||||
6503 TCCCTCCGCGCAGGAGGA.....AACGAGCAG..... 6475
  |||||
211 LeuGlnTrpThrProPheSerSerAlaAspLeuTyrAsnTrpLysThr 227
  |||||
6474 CTCAATACAGCGCGTCTCTCTCTGACCTTTACACTGCAAAATTA 6425
  |||||
227 nHisProProPheSerGluAspProGlnArgLeuThrGlyLeuValGlu 244
  |||||
6424 TAACCTCTCTTTTCTGAAGATCCAGTAAACAGTCACTGATGAGCT 6375
  |||||
244 eLeuMetPheSerHisGlnProThrTrpAspAspCysGlnLeuLeu 260
  |||||
6374 CTGTTCTCATACCCATCCAGCCCTGGAGAGACTGTCAGAGCTGTTG 6325
  |||||
261 GlnThrLeuPheThrThrGluArgGluArgGlyLeuLeuGlnAla 277
  |||||
6324 GGGACTCTGTGACCGGAGAGAAACAAACGAGGCTCTTGAAGGCTAG 6275
  |||||
277 gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGlu 294
  |||||
6274 AAGGGGAGTGGGGGAGATGATGGGCGCCCACTCAACGCCAATGAAG 6225
  |||||
294 LeaspMetGlyPheProLeuThrArgProGlyTrpAspTyrAsnThr 310
  |||||
6224 TCGATCCGCTTTTCCCTCGAGCGCCAGACGAGGATTACACACCCAG 6175
  |||||
311 GluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGly 327
  |||||
6174 GCAAGTGAAGAACCACTAGTCCACTATCCGCACTTCTCTACCGGCTT 6125
  |||||
327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluVal 344
  |||||
6124 CCAAAACGGGGGAGAAAGCCCAATTTGGCAAGTAAGAAAGATTA 6075
  |||||
344 eGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
  |||||
6074 CACAAAGGGCCCAATGAGTCTCCCTCGCTTCTTCAAGAGACTTAAGA 6025
  |||||
361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLys 377
  |||||
6024 GCCTATCGCAGTACCTCTTATGACCTGAGGAGACCCAGGAGAAAC 5975

```



```

377 aserValAlaLeuAlaPheIleGlyInserAlaLeuAspIleArgLysL 394
5974 TATATGCTATATGCTTTTCATTGGCAGCTCCCGACATTTGGAGAGA 5925
394 yslLeuGlnArgLeuGlnGlyLeuGlnGlnAlaGluLeuArgAspLeuVal 410
5924 AGTTAGAGAGAGTTAGAGATTAAAAACACACACCTTGCAGATTGGTT 5875
411 ArgGluAlaGluLysValTyrTyrArgArgGluThrGluGluLysGlu 427
5874 AGAGAGGCGAAGAAAGATCTTTAATAACGAGAAACCCCGAAGAAAGAGA 5825
427 uGlnArgLysGluLysGluArgGluLysGluLysGluLysGluLys 440
5824 GGAACCTATACGAGAGAGAAACAGAGATCGTAGAGAGACATAGAGAGATG 5775
441 .....ArgAspArgArgGlnGlnLysAsnLeu 449
5774 AGCATGAGCAGAAAGAAAGAAAGATCGTAGAGAGACATAGAGAGATG 5725
450 ThrLysIleLeuAlaValAlaValGluLysSerSerArgGluArgGlu 466
5724 AGCAAGCTATTGGCCACTGCTGTAGTGACAGAAACAG ..... 5686
466 uArgSpheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuG 483
5685 .....GARGAGACAGGAGAGAGAA...C 5667
483 IYAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLys 499
5666 GAAAGAGGTCCCACTCGATCCGACGACAGTGTGCTACTGCAAAAGAAAAG 5617
500 GlyHisThrPalaArgAsnCysProLysLys.....GlyAsnLysGlyPr 514
5616 GGGGCACTGGGCTAAGATTGTCCCAAGAAACACAGAGACCTCGGGGACC 5567
514 Olys.....ValLeuAlaLeuGlnGlu 521
5566 AAGACCCAGACACTCTCTGACCTGAGCTAGATGAC 5533
seq_name: /cgn2_6/prodata/2/1na/3A_COMB.seq:US-08-110-300A-8
seq documentation block:
; Sequence 8, Application US/08110300A
; Patent No. 5643736
; GENERAL INFORMATION:
; APPLICANT: Plinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull and Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110.300A
; FILING DATE: 20-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H.
; REGISTRATION NUMBER: 32,140
; REFERENCE/DOCKET NUMBER: 11698A50
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461

```

```

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-110-300A-8

alignment_scores:
    Quality: 1350.50      Length: 589
    Ratio: 3.535          Gaps: 14
    Percent Similarity: 64.856      Percent Identity: 47.538

alignment_block:
US-09-171-553B-4 x US-08-110-300A-8 ..

Align seq 1/1 to: US-08-110-300A-8 from: 1 to: 8323

1 MetGlyInThrValThrProLeuSerLeuThrLeuAspHisThrph 17
619 ATGGGCGAGGCTGTACACCCCTTAAGTTGACTTACACCACTGAGAA 668
17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
669 GGATGTGAAAGGACAGACCCACAACTGTGTGAGAGTTAGAAAAGGC 718
34 rOTrgInThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
719 GTGGGTTACATTCTGCTGTGAGAAATGGCCAACTTCAACGTGCGATG 768
51 ProSerGluLysThrPheAsnSerGluIleIleLeuAlaValLysAla 67
769 CCAAGAGCGGACATTAAACCCAGCAATTATTACAGAGTTAAGATCAA 818
67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
819 GGTCTTCTACACCTGGCCACATGACATCGGATGACGTCCCTTACATG 868
84 euthTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
869 TGACCTGGGAAGCTATAGCAGTAGACCCCTGCTGGGTCAGACCTTC 918
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGlu 117
919 GTGCAC...CTTAACCTCC...CTCTCTT ..... 945
117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuPro 134
946 .....CCCCCTGACGCCCTCTCTCCAC 970
134 rGAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
971 CTGAA.....CCCCACTCTGACCCCG 993
151 ProLeuSerSerThrGlyCysGluGlyThrSerAlaProProGlyAla 167
994 CCCCAGCTCTCCCTC.....TA 1010
167 aProValAlaGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 184
1011 TCCGGCTCTCACTTCTCT...TTAAACCAACCAACTAGGCTCAAGTCC 1057
184 hProGluArgThrAspGluIleAlaIleLeu..... 194
1058 TTCCTGATAGCGAGAGACCATCATGATCTACTCAGGAGAGACCTCCG 1107
195 ProLeuArgThrTyrGlyProProMetProGlyGly..... 206
1108 CCTTACCGGGACCCAGGGCCACCTCTCTGACGGAACGGCGATAGCGG 1157

```

```

206 ..... 206
1158 AGAAGTGGCCCTACAGAGAGCCCTGACCTTCCCAATGATATCCC 1207
207 .....Gln 207
1208 GCCTGGCGGGAAGAAAGAACCCCTGGCGGATTTACTACTACTCCAG 1257
208 LeuGlnProLeu.....GlnTyrTrpProh 216
1258 GCGTTCCCTTCGCTGGAGGAGATGACATATCATATCGCCATT 1307
216 eSerSerAlaSplLeuTyrAsnTyrPylsThrAsnHisProProhSerG 233
1308 TTCCTCTCTGACCTCTATTAACGAAATACAGACCCCTCTTCTCCG 1357
233 LuAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHis 249
1358 AGGACCCAGCTAATTGACAGCTTGATCGATCGCTCTCTACTCAT 1407
250 GlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThr 266
1408 CAGCCCACTTGGATGACGACCAACAGCTATTAGGACCTGCTGACGG 1457
266 rGluGluArgGluArgIleLeuLeuGlnAlaArgLysAsnValProGly 283
1458 AGAAGAAACACAGCGAGTCTCTAGAGGCCCAAGCGGTTGACGAGG 1507
283 LaAspGlyArgProThrGlnLeuGlnAsnGlnIleAspMetGlyPhePro 299
1508 AGACGAGACCCCACTGACCTGCCCAAGACATTAAATGATGCTTTCC 1557
300 LeuThrArgProGlyTyrAspTyrAsnThrAlaGluArgGluSerLe 316
1558 TTGGAGGTCCGACGCTGGACTCAACACCAAGAGTACGAGACACCT 1607
316 uLysIleTyrArgIleAlaLeuValAlaGlyLeuArgIleGlyAlaSerAla 333
1608 AGTCCACTATTCGACAGTTCCTCAGCGTCTCCCAAAAGCGGAGAA 1657
333 rGpProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGlu 349
1658 GCCCACCACATTGGCCAGGTAAAGAGGATTAACCCAGGACCTATGAG 1707
350 ProProSerValPheLeuGluArgLeuMetGluAlaPheArgPheTh 366
1708 TCTCCCTCAGCGCTTTTAGAGAGACTCAAGAGGCTATCGCAGATAC 1757
366 rProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaP 383
1758 TCCTTATGACCTGAGGAGCCAGGCAAGAAACCAATAGTGGCATGTGCAT 1807
383 heIleGlyIleSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
1808 TCATGTGACATTCGCGCCCGGATATCGGCGCAATAGAGCGTTAGAA 1857
400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysVa 416
1858 GATTGAGAGAGTAAGACCTTAGGAGACTTATGAGGAGAGCTGAAAGAT 1907
416 lTyrTyrArgArgGluThrGluGluGluLysGluGlnArgLysLysG 433
1908 CTTTATATAAGAGAAACCCCGGAAGAAAGAGAAAGTATATGAGAGAG 1957
433 LuArgGluGluArgGluArg..... 440
1958 AAACAGAGGAAAGAAAGAACCCCTAGGCGAGAGATGTCCAGAGAGAG 2007
441 .....ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAl 455
2008 AAGGAGAGGAGCCGAGAAAGACATAGAGAAATGACTTAAGTTGCTGCTAC 2057

```

```

455 aValValGluGlyLysSerSerArgGluArgGluArgAspPheArgLysI 472
2058 TGCTGTTAGCCGCGCAG.....AGACAGG 2080
472 leArgSerGlyProArgGlnSerGlyAsnLeuGlnLysAsnArgThrProLeu 488
2081 ATAGACAGGAGAGAGAGCA.....AGAGGCCCAACATC 2115
489 AspLysAspGlnCysAlaTyrCysLysGluLysGlyHisThrAlaArgAs 505
2116 GACCCAGCAGCAGTGTGCTACTGCTCAAGAAAGAGACATTGGCTAGAGA 2165
505 ncysProLys.....GlyAsnLysGlyProLys.....V 516
2166 TTGCCCAAGAACCAAGAGACCCCGGAGACACAGCCAGCCGCTGCC 2215
516 aLeuAlaLeuGluGlu 521
2216 TCCTGACCTTAGACGAT 2232

seq_name: /cogn2_6/ptodata/2/ina/5R_COMB.seq:us-08-886-642-8

seq_documentation_block:
? Sequence 8, Application US/0886642
? Patent No. 5952474
? GENERAL INFORMATION:
? APPLICANT: Pinter, Abraham
? APPLICANT: Kayman, Samuel
? TITLE OF INVENTION: FUSION GLYCOPROTEINS
? NUMBER OF SEQUENCES: 25
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson P.C.
? STREET: 45 Rockefeller Plaza, Suite 2800
? CITY: New York
? STATE: NY
? COUNTRY: USA
? ZIP: 10111
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows95
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/886,642
? FILING DATE: 01-JUL-1997
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/110,300
? FILING DATE: 20-AUG-1993
? APPLICATION NUMBER: 07/938,100
? FILING DATE: 28-AUG-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Hone, William J.
? REGISTRATION NUMBER: 26,739
? REFERENCE/DOCKET NUMBER: 07763/010002
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212/765-5070
? TELEFAX: 212/258-2291
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8323 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: Genomic DNA
? US-08-886-642-8

alignment_scores:
? Quality: 1350.50 Length: 589
? Ratio: 3.535 Gaps: 14
? Percent Similarity: 64.856 Percent Identity: 47.538

alignment_block:

```

US-09-171-553B-4 x US-08-886-642-8 ..

Align seg 1/1 to: US-08-886-642-8 from: 1 to: 8323

```

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuSphistPth 17
  |||||
619 ATGGGCCAGGCGTGTACCCCTTAAGTTGACTTTAGACCACTGGAA 668
  |||||
17 rGluValArgSerArgAlaHisAsnLeuSerValGluValLysGlyP 34
  |||||
669 GGATGTGAGACGAGACGCCAACACGTCGTGAGAGTTAGAAAGGC 718
  |||||
34 rOTpGlnThrPheCysAlaSerGluThrProThrPheAspValGlyTrp 50
  |||||
719 GCTGGTTACATCTCTGCTGCAAAATGGCCAACTTCACATCGGATGG 768
  |||||
51 ProSerGlnGlyThrPheAsnSerGluLeuLeuAlaValLysAla 67
  |||||
769 CCAGGAGACGCGACTTTTAAACCGACATTTACACAGGTTAGATCAA 818
  |||||
67 eLlePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
  |||||
819 GGTGTTCTACCTGGCCACATGACATCGGATCGATCCCTCATATCG 868
  |||||
84 eutHTrpGlnAspLeuAlaGluAspProProProProValLysProTrp 100
  |||||
869 TGACCTGGAGAGCTATAGCAGTAGACCCCTCCGCTGAGACCTTC 918
  |||||
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGly 117
  |||||
919 GTGCAC...CTTAAACCTCC...CTCTGCTT... 945
  |||||
117 sAsnLysHisSerAlaGluValGluProSerSerSerTyrIleProA 134
  |||||
946 .....CCCCCTGACGCCCTCTCTCCAC 970
  |||||
134 rGAspArgLysAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
  |||||
971 CTGAA.....CCCCACTCTCAACCCC 993
  |||||
151 ProLeuSerSerThrGlyCysCysGlnGlyThrSerAlaProProGlyAl 167
  |||||
994 CCCCACTCTCCCTC.....TA 1010
  |||||
167 aProValAlaGlnGlyProAlaAlaGlyThrArgSerArgArgLysAla 184
  |||||
1011 TCCGGCTCTCACTCTCTCT...TTAAACCAACCACTAGGCTTAAGTCC 1057
  |||||
184 hProGluArgThrAspGluIleAlaIleLeu..... 194
  |||||
1058 TTCCGATAGCGGAGAGACCACTCATTTGATCTACAGGAGACCTTCG 1107
  |||||
195 ProLeuArgThrTyrGlyProPheProGlyGly..... 206
  |||||
1108 CCTTACCGGAGACCGGACCCCTCTCTGAGGAGGAGCGCATATCGG 1157
  |||||
206 ..... 206
  |||||
1158 AGAAGTGCCCTACAGAGAGAGCCCTGACCTCCCAATGTAATCC 1207
  |||||
207 .....Gln 207
  |||||
1208 GCTGCGGGGAGAAAGAACCCCGGCGGATTCTACATCTCTTCG 1257
  |||||
208 LeuGlnProLeu.....GlnTyrTrpProh 216
  |||||
1258 GCGTTCCTCCCTGCGGAGGAGATGACATCATCTACTGCGCATTT 1307
  |||||
216 eSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProPheSerG 233
  |||||
1308 TTCTCTCTCTGACCTCTATTAACGGAATAACAAACCCCTCTTCTCCG 1357
  |||||
233 LuAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHis 249
  |||||
1358 AGGACCCAGCTAAATTTGACAGCTTTGATCGAGTCCGTTCTTACTCAT 1407
  |||||
250 GlnProThrTrpPaspAspCysGlnGlnLeuGlnThrLeuPheThr 266
  |||||
1408 CAGCCACTTGGATGATGCTGCAACACCTATTAGAGGACCTGCTGACGG 1457
  |||||
266 rGlnGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyA 283
  |||||
1458 AGAAGAAAAACAGGAGATGCTCTAGAGGCCCAAGAGGCGGTGAGGGG 1507
  |||||
283 laAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPhePro 299
  |||||
1508 AGGAGGAGACCCCACTCAGCTCCCAATGATCATTAATGATGCTTTCC 1557
  |||||
300 LeuThrArgProGlyTrpAspTyrAsnThrAlaGlnGlyArgGluSer 316
  |||||
1558 TTGGAAAGTCCCGACTGGAGCTGACCAACACCAAGAGGATGAGACCAT 1607
  |||||
316 uLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArg 333
  |||||
1608 AGTCACATATCGCCAGTCTCTAGCGGCTCCAAAAGCGGCGAGAA 1657
  |||||
333 rProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGlu 349
  |||||
1658 GCCCACAACATTTGGCCAGATGAAGATGAACCCAGGAGCTTAATGAG 1707
  |||||
350 ProProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheTh 366
  |||||
1708 TCTCCCTCAGCCTTTTAAAGAGACTCAAGAGGCTTTCAGATACAC 1757
  |||||
366 rProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAla 383
  |||||
1758 TCCTTATGACCTGAGAGACCGGCAAGAAACATGCGCATATGCAT 1807
  |||||
383 heLleGlyLysSerAlaLeuAspIleArgLysLysGlnArgLeuGlu 399
  |||||
1808 TCATCTGCGAGTCCGCCGATATCGGCGCAAAAGTTAGAGGTTAGAA 1857
  |||||
400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLys 416
  |||||
1858 GATTTGAAGAGTAAGACCTTAGAGACTTAGAGGAGCACTGAAGAT 1907
  |||||
416 lTyrTyrArgArgGluThrGlnGluGlyGlnArgLysGluLysG 433
  |||||
1908 CTTTATATGAAGAAACCCCGGAAAGAAAGAGAGGATTTAGGAGAG 1957
  |||||
433 LuArgGlnGluArgGluGluArg..... 440
  |||||
1958 AAACAGAGAGAAAGAAAGACCGGTAGGCGACATGTGACAGAGAG 2007
  |||||
441 .....ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAla 455
  |||||
2008 AAGGAGAGGAGACCGCAACAGATAGAAATAGTAATGCTGCTGCTAC 2057
  |||||
455 aValValGluGlyLysSerSerArgGluArgGluAspPheArgLysI 472
  |||||
2058 TGTGTTAGGGGGCG.....AGACAG 2080
  |||||
472 leArgSerGlyProArgGlnSerGlyAsnLeuGlnLysAsnArgThrPro 488
  |||||
2081 ATAGACAGGAGAGAGAGCA.....AGAGGCCCCCACTC 2115
  |||||
489 AspLysAspGlnCysAlaTyrCysLysGluLysGlnHisTrpAlaArg 505
  |||||
2116 GACCAGAGACAGTGTCTACTAGCAAGAAAGGACATTTGGCTAGAGA 2165
  |||||
505 nCysProLysLys.....GlyAsnLysGlyProLys.....V 516
  |||||
2166 TTGCCCCAAGAAAGCAAGAGAGACCCGGGACCAAGACCCAGGCTTCC 2215
  |||||
516 aLeuAlaLeuGluGlu 521
  |||||

```

2216 TCCTGACTTGAACGAT 2232

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-8

seq_documentation_block:

```

Sequence 8, Application PC/TUS9308041
GENERAL INFORMATION:
APPLICANT: The Public Health Research Institute of the City of
APPLICANT: New York, Inc.
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESS: Davis Hoxie Faithfull and Hapgood
STREET: 45 Rockefeller Pl.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08041
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,100
FILING DATE: 28-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/____
FILING DATE: 20-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jacobs, Seth H.
REGISTRATION NUMBER: 32,140
REFERENCE/DOCKET NUMBER: 11698A50
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-757-2200
TELEFAX: 212-586-1461
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 8323 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
PCT-US93-08041-8

```

alignment_scores:

```

Quality: 1350.50      Length: 589
Ratio: 3.535          Gaps: 14
Percent Similarity: 64.856      Percent Identity: 47.538

```

alignment_block:

US-09-171-553B-4 x PCT-US93-08041-8 ..

Align seg 1/1 to: PCT-US93-08041-8 from: 1 to: 8323

```

1 MetGlnThrValThrProLeuSerLeuThrLeuAspHisTrpTh 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
619 ATGGGCGAGCGCTGTACACCCCTTAAGTTGACTTAGACCACTGAA 668
17 rglValArgSerArgAlaHisAsnLeuSerValGlnValIysLysGly 34
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
669 GGAATGTCAGACGACGCCCAACCTGTGCTGAGAGTTAGAAAAAGGC 718
34 rorTpgInThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
||| |||||:|||||:|||||:|||||:|||||:|||||:
719 GCTGGTACATCTGCTCTGCAAGATGCGCAACCTTCAACGTGGATGG 768

```

```

51 ProSerGlnGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIle 67
||| :|||||:|||||:|||||:|||||:|||||:|||||:
769 CCACGAGAGCGGCGACTTTTACCCAGACATTTATACACAGCTTAGATCAA 818
67 eilePheGlnThrGlyProGlySerHisProAspGlnIleProTyrIleIle 84
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
819 GGTCTTTCACCTGGCCCACTGACATGACATTCGGATCAGTCCCTACATCG 868
84 eutThrTrpGlnAspLeuAlaGluAspProProProTyrValLysProTyr 100
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
869 TGACCTGGAGAGCTATAGACAGTACACCCCTCCCTGGTGCAGACCTTCG 918
101 LeuAsnLysProArgLysProGlyProArgTleLeuAlaLeuGlyLysIly 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
919 GTGCAC...CCTAATCCTCC...CCTCTCTT... 945
117 sAsnLysHisSerAlaGluLysValGluProSerSerTyrLeuProAla 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 .....CCCCCTTACGCCCTCTCTCCAC 970
134 rgaAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
971 CTGAA.....CCCCCACTCTGACCCCG 993
151 ProLeuSerSerThrGlyCysGlyGlyThrSerAlaProProGlyAla 167
||| |||||
994 CCCAGTCTCTCCCTC.....TA 1010
167 aproValAlaGlnGlyProAlaAlaGlyThrArgSerArgGlyAlaTr 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1011 TCGGCTCTCAGTCTCTCT...TTAAACCAACCAACCTAGGCCCTAGTCC 1057
184 hrProGluArgThrAspGlnIleAlaIleLeu..... 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1058 TTCCTGATACGGAGGACACCATATGATCTACTACGAGACACCTCCG 1107
195 ProLeuArgThrGlyProPheProMetProGlyGly..... 206
||| ||| |||||:|||||:|||||:|||||:|||||:
1108 CTTTACCGGAGCCAGGCGCCCTCTCTGACGGAGAACGGCGATAGCGG 1157
206 ..... 206
1158 AGAAGTGGCCCTACAGAGAGAGCCCTTCCCAATGATATCC 1207
207 .....Gln 207
1208 GCCTGCGGGAGAGAAAGAACCCCGTGGGATTTCTACTACCTCTCAG 1257
208 LeuGlnProLeu.....GlnTyrTrpProPh 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1258 GCGTTCCTCCCTTCCGCTGGGAGGGAATGACAGTATCATATACATGCGCAT 1307
216 eSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerG 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1308 TTCTCTCTGACCTGTATTAATGGAATAATACAAACCCCTCTTCTCGG 1357
233 LAspProGlnAlaGlyLeuThrGlyLeuValGlnSerLeuMetPheSerHis 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1358 AGGACCCAGCTAAATTGACAGCTTGTATGAGAGCTCTCTCTTACTCAT 1407
250 GlnProThrTrpAspAspCysGlnGlnIleLeuGlnThrLeuPheThrTh 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1408 CACCCCACTTGGGATGACTGCCAACAGCTATTATGAGACCCCTGCTGAGGG 1457
266 rglGluArgGluArgIleLeuLeuGlnAlaArgLysAsnValProGlyAla 283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1458 AGAAGAAAAACAGCGAGTCTCTTACAGGCCGGAAGCGGTTGCGAGGG 1507
283 LAspGlyArgProThrGlnLeuGlnAsnGlnIleAspMetIlePhePro 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1508 AGGACGAGAGCCCACTCAGCTGCCAATGACATTAATGATGCTTTTCC 1557

```

```

300 LeuThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgIuSerIe 316
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1558 TTGAGACGTCCCGACTGAGTACACACCCAGAGTAGAGACCACT 1607
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 ULSTLeYrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArg 333
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1608 AGTCACATATGCGAGTGTCTCCAGGGGTCTCCAAACCGGGGAGAA 1657
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
333 rGpThrAsnLeuAlaValArgGluValMetGlnGlyProAsnGlu 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1658 GCCCACCACATTTGGCCAGGTAAAGGATACCCAGGAGCCTATAGAG 1707
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 ProProSerValPheLeuGluArgLeuMetGlnAlaPheArgArgPhe 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1708 TCCTTATACCTTTTGTAGAGACTCAAGAGCGCTATGCGAGATACAC 1757
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 rProPheAspProThrSerGluAlaGlnGlyAlaSerValAlaLeuAla 383
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1758 TCCTTATACCTTTGAGAGCCAGGCAAGAACCAATGTGGCCATGTGAT 1807
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
383 heileGlyInSerAlaLeuAspIleArgGlyLysLeuGlnArgLeuGlu 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1808 TCATCTGGCAGTCCGCCCCGATATCCGCGCAAGATTAGAGCGGTGAA 1857
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGlyIu 416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1858 GATTGAGAGTAAAGCCTTGAAGACTTACTGAGGAGAAAGTGAAGAT 1907
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 ITyTrArgArgGluThrGlnGluGlyLysGlnGlyArgGlyLysGlu 433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1908 CTTTATATTAACGAGAACCCCGGAGAAAGAGAGAACGATTTAGAGAG 1957
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 IuArgGluGluArgGluArgGluArgGluArgGluArgGluArgGlu 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1958 AAACAGAGAGAAAGAGAACCCGCTAGGCGAGAGATGTGCAGAGAGAG 2007
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
441 .....ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAla 455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2008 AAGGAGAGGAGACCGCAGAACATAGAGAAATGAGTGAAGTGGCGGTAC 2057
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
455 aValValGluGlyLysSerSerArgGluArgGluArgGluArgGlu 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2058 TGTCCTTACCGCGCAG.....AGACAGG 2080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
472 leArgSerGlyProArgGlnSerGlyAsnLeuGlyAsnArgThrPro 488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2081 ATAGACAGAGAGAGAGAGCA.....AGAGGCGCCCAACTC 2115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
489 AspLysAspGlnGlyAlaTyrCysGlyGluGlyLysIleThrPalaArg 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2116 GACCAAGCAGAGTGTGCTACTGCAAGAAAGGAGACATTTGGGCTAGAGA 2165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
505 nCysProLysLys.....GlyAsnLysGlyProLys.....V 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2166 TTGGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2215
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 alleuAlaLeuGluGlu 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2216 TCCTGACCTTAGACGAT 2232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: /cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-110-300A-9
seq_documentation_block:
; Sequence 9, Application US/08110300A
; Patent No. 5643756
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull and Hapgood
; STREET: 45 Rockefeller Pl.

```

```

? CITY: New York
? STATE: New York
? COUNTRY: USA
? ZIP: 10111
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/110,300A
? FILING DATE: 20-AUG-1993
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Jacobs, Seth H.
? REGISTRATION NUMBER: 32,140
? REFERENCE/DOCKET NUMBER: 11698A50
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-757-2200
? TELEFAX: 212-586-1461
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 10367 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? US-08-110-300A-9

```

```

alignment_scores:
    Quality: 1350.50      Length: 589
    Ratio: 3.535          Gaps: 14
    Percent Similarity: 64.856      Percent Identity: 47.538

```

alignment_block:

US-09-171-553b-4 x US-08-110-300A-9/rev ..

```

Align seg 1/1 to reverse of: US-08-110-300A-9 from: 1 to: 10367
1 MetGlyGlnThrValThrThrProLeuSerIleuThrLeuAspHisThr 17
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6977 ATGGGCGAGGCGTGTACCAACCCCTTAAGTTGACTTAGACCAAGGAA 6928
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysGly 34
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6927 GGAATGCGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6878
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 rotPrGlnThrPheGlyAlaSerGluTrpProThrPheAspValGlyTrp 50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6877 GCGGGGTTCATCTCTCTGCAAGTGGCCCAACCTTCAACGTCGATGG 6828
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ProSerGlnGlyThrPheAsnSerGluIleIleLeuAlaValLysAla 67
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6827 CCACGAGAGCGCACTTTTAAACCGACGATTTATACAGAGTTAAAGTCA 6778
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 eilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6777 GGTCTTCTACCTGCGCCACATGAGACATCGGATGAGTCAAGTCCATAG 6728
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 eutThrPrGlnAspLeuAlaGluAspProProThrValLysProTrp 100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6727 TGACCTGGGAAGCTATAGACAGTACCCCTCCCTGGGCTAGAGACCTTC 6678
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGly 117
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6677 GTGCAC...CTTAACCTCC.....CTCTCTCT..... 6651
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 sAsnLysHisSerAlaGluValGluProSerSerSerTyrLeuPro 134
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6650 .....CCCCCTTACGCCCCCTCTCTCCAC 6626

```


REFERENCE/DOCKET NUMBER: 07763/010002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212/765-5070
 TELEFAX: 212/258-2291
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10367 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 US-08-886-642-9

alignment_scores:
 Quality: 1350.50 Length: 589
 Ratio: 3.535 Gaps: 14
 Percent Similarity: 64.856 Percent Identity: 47.538

alignment_block:

US-09-171-553b-4 x US-08-886-642-9/rev ..

Align seg 1/1 to reverse of: US-08-886-642-9 from: 1 to: 10367

1 MetGlyInThrValThrProLeuSerLeuThrLeuAspHisTyrPth 17
 |||||
 6977 ATGGGCGAGGCTGTACCCCTTAAGTTGACTTTAGACCTGGAA 6928
 17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
 |||||
 6927 GGATGTGACGACGACGACGACCAACCTGTCGTAGAGGTAGAAAGGC 6878
 34 rGTrGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
 |||||
 6877 GCTGGGTTACATTCTGCTCTGCGAATGGCCACCTTCAACGTGGAGTG 6828
 51 ProSerGluGlyThrPheAsnSerGluLeuLeuAlaValLysAla 67
 |||||
 6827 CCACAGACGACGACTTTAACCCAGACATTATTACACAGGTAAAGATCA 6778
 67 eLeuPheLeuThrGlyProGlySerHisProAspGlnGluProTyrLeu 84
 |||||
 6777 GGTCTTCACCTGGCCGACATCCGATCCGATCCGATCCGATCCGATCCG 6728
 84 eutHrTrpGlnAspLeuAlaGluAspProProProTyrValLysProTyr 100
 |||||
 6727 TGACCTGGAGACTATAGACATAGACCCCTCCCTGGGTCAGACCTTC 6678
 101 LeuAsnLysProArgLysProGlyProArgLLeuAlaLeuGlyGly 117
 |||||
 6677 GTGCAC...CCTAAACCTCC...CTCTCTT... 6651
 117 sAsnLysHisSerAlaGluLysValGluProSerSerTyrLeuPro 134
 |||||
 6650CCCCCTTCAGCCCTCTCTCCAC 6626
 134 rGaAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
 |||||
 6625 CTGA... ..CCCCCATCTCTGACCCCG 6603
 151 ProLeuSerSerThrGlyCysGlyGlyThrSerAlaProProGlyAl 167
 |||||
 6602 CCCAGTCTCTCC... ..TA 6586
 167 aProValValGluGlyProAlaAlaGlyThrArgSerArgArgLysAla 184
 |||||
 6585 TCCGGCTCTCACTTCTCT...TTAAACAACAACCTAGGCTCAAGTCC 6539
 184 hrProGluArgThrAspGluLeuAlaLeu... .. 194
 |||||
 6538 TTCTGTATGAGGAGGACACCTATTGATCTACTACGAGAGACCTCCG 6489
 195 ProLeuArgThrTyrGlyProPrometProGlyGly... .. 206

6488 CCTTACCGGACCCAGGCGCACCTCTCTACCGGAACGGCGATAGCG 6439
 206 206
 6438 AGAAGTGGCCCTACAGAGAGCCCTGACCTTCCCAATGATGCC 6389
 207Gln 207
 6388 GCCTGGGGGAAGAAAGAACCCCGTGGCGGATTTCTACTACCTCAG 6339
 208 LeuGlnProLeu.....GlnTyrTrpPro 216
 6338 GCGTTCCTCCCTTCCGCGGAGGAGGATGACAGTATCAATGCGCAT 6289
 216 eSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSer 233
 6288 TTCTCTCTGACCTCTATAGCTGGAATAATACACCCCTTTCTCCG 6239
 233 LuAspProGlnArgLeuThrGlyLeuValGlnSerLeuMetPheSerHis 249
 6238 AGGACCCAGCTAAATTGACAGCTTGTATGAGTCCGTTCTCTACTCAT 6189
 250 GlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThr 266
 6188 CAGCCCACTGGGATGATGACTCCCAACAGCTATTAGGACCTCTGACGG 6139
 266 rGluGluArgGluArgGlyLeuLeuGlnValArgLysAsnValProGly 283
 6138 AGAAGAAACACAGGAGTCTCTAGAGGCCCAAGGCGGTTGAGGGG 6089
 283 LuAspGlyArgProThrGlnLeuGlnAsnGlnLeuAspMetGlyPhePro 299
 6088 AGGACGGACCCCACTACCTGCCAATGACATTATGATGCTTTCC 6039
 300 LeuThrArgProGlyTyrAspTyrAsnThrAlaGlyLysGlySerLe 316
 6038 TTGGAAGTCTCCGACGTGGGACTTACACACCCCAAGGATGAGAACCT 5989
 316 uLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgLysAlaSerArg 333
 5988 AGTCCACTATGCGGACGATGCTCTAGCGGCTCTCCAAAGGGGAGCA 5939
 333 rGProThrAsnLeuAlaLysValArgGluValMetGlnLysProAsnGlu 349
 5938 GCCCACAATTTGGCCAGTAAGATTAACCCAGGACTATATAG 5889
 350 ProProSerValPheLeuGluArgLeuMetGluAlaPheArgPheThr 366
 5888 TCTCCCTGACCTTTTAAAGAGACTCAAGGAGGCTATCCGAGATAC 5839
 366 rProPheAspProThrSerGluAlaGlnLysLysSerValAlaLeuAla 383
 5838 TCCTTATGACCTGAGGACCCAGGCAAGAAACATGATGCGCATGTAT 5789
 383 heLeuGlyInSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
 5788 TCATCTGACATGCGCCCGGATATCGGGCAAACTTAGAGCGTTAA 5739
 400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLys 416
 5738 GATTTGAAGAGTAAAGACCTTAGAGACTTAGAGAGGAACTGAAGAT 5689
 416 LysTyrArgArgGluThrGluGlnLysGlnGlnArgLysGlyLysG 433
 5688 CTTTATTAACAGAGAAACCCGGAAGAAAGAGAGATTTAGAGAG 5639
 433 LuArgGluGluArgGluGluArg..... 440
 5638 AAACAGAGAAAGAGAGACCGGTAGGCAAGAGATGTCCAGAGAG 5589
 441ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAla 455


```

5588 AAGGAGGGGACCGCAGACATAGAGAAATGAGTAAGTTGCTGCTAC 5539
455 aValValGIuGIyLysSerSerArgGIuArgGIuArgAspHeArgLysI 472
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5538 TGTCGTTAGCGGCGCAG.....AGACAGG 5516
472 leArgSerGIyProArgGIuSerGIyAsnLeuGIyAsnArgThrProLeu 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5515 ATAGACAGGGAGGAGAGCGA.....AGGAGGCCCACTC 5481
489 AspLysAspGIuGIyAlaTyrCysLysGIuLysGIyHisTPAlaArgAs 505
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5480 GACCAAGACAGCGTGTGCTACTGCAAGAAAGGAGACATTGGGCTAGAGA 5431
505 nCysProLysLys.....GIyAsnLysGIyProLys.....V 516
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5430 TTGCCCCAAGAGCCAGAGAGACCCCGGAGACGACCCCAAGCCTCCC 5381
516 aLLeuAlaLeuGIuGIu 521
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5380 TCCTGACCTTAGACGAT 5364

```


2210 GGAGGGGCAACCACTTGATGTCCGTGTTGATACCAGGAGCGAACAATTACG 2259

34 allLeuleugInProLeugLYLysLeulYsgLylsYssertYPvalMet 50
|||||
|||:::
2260 TGCTACTACACGCCATTAGGAATAAAAGATAAAAATCCTGGTGGATG 2309

```

51 GlyAlaThrGlnArgGlnTyrProThrThrArgArgThrValAs 67
|||||
2310 GGTGCCACAGGGGACACACATATCATCATGACATCCGAGAACAGTGTGA 2359
67 PheGlnValAlaArgValThrHisSerPheLeuValIleProGluCysP 84
|||||
2360 CTTGGGAGTGGGACGGGTAAACCACTCTCTCTGTCATACCTGAGTGC 2409
84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
|||||
2410 CAGCACCCCTCTTAGGTAGAGACTATTATGACCAGATGGGACCAAT 2459
101 SerPheGlnGlnGlyArgProGlnValSerValAsnAsnLysProIleTh 117
|||||
2460 TCTTTGAAACAAGGGAACACAGAACTGTCTGCAATAACAACCTATCAC 2509
117 rValLeuThrLeuGlnLeuAspAsnGlnTyrArgLeuTyrSerProGln 134
|||||
2510 TGTGTGACCTCTCCAAATTAGATGACCAATATCGACTATCTCTCCCTAG 2559
134 alLysProAspGlnAspIleGlnSerTyrLeuGlnGlnPheProGlnAla 150
|||||
2560 TAAAGCTGATCAAAATATCAATCTGTTGGAACAGTTCGCCCAAGCC 2609
151 TrpAlaGlnThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVa 167
|||||
2610 TGGGCAGAAACCCGACGAGATGGGTTGGCAAGCAAGTTCGCCCAAGT 2659
167 lIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrProL 184
|||||
2660 TATTCAACTGAAGGCCAGTCCACACACACATGTCAGTCAAGACAGTACCCT 2709
184 euSerArgGlnAlaArgGlnGlyIleTrpProHisValGlnArgLeuIle 200
|||||
2710 TGAGTAAAGAGCTCAAGAGAGATTCGGCCGATGTCCTCAAGATTAAATC 2759
201 GlnGlnGlyIleLeuValProValGlnSerProTyrPAsnThrProLeu 217
|||||
2760 CAACAGGGGCAATCCTAGTCTCTGTCATCTCCCTCGAATACATCCCTGCT 2809
217 upProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeu 234
|||||
2810 ACCGGTAAAGAAAGCTGGGACTATGACTATCAACACAGTACAGAGACTTGA 2859
234 rGlnValAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
|||||
2860 GAGAGGTCATTAAGGGGTGCGAGATATACACCAACAGTCCGGAACCT 2909
251 TyrAsnLeuLeuSerAlaLeuProProGlnArgAsnTrpTyrThrVal 267
|||||
2910 TATAACCTCTGTGTGCTCTGCCACCCCAAGGAGTGTATACAGTAT 2959
267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
|||||
2960 GGACTTAAAGATGCTCTTCTCTGCTGAGATTACACCCACACAGTACCAAC 3009
284 roLeuPheAlaPheGlnTyrArgAspProGlyThrGlyArgThrGln 300
|||||
3010 CACTTTTGGCTCTCAATGAGATCCAGGTACGGAAGAACCGGAG 3059
301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePh 317
|||||
3060 CTCACCTGGACCCGACTCCCAAGGGTTCAAGAACTCCCGGACCACTTT 3109
317 eaSpGlnAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
|||||
3110 TCACCAAGGCTTACACAGAGACTGGCCCACTTCAAGGATCCACACACCTC 3159
334 lValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAlaThr 350
|||||
3160 AGGTGACCTCTCTCCAGTACGTGATGACCTGCTTGGCGGAGCACC 3209
351 LysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuGlnLeuSerAs 367
|||||
3210 AAACGAGACTCTTAGAAGGACACAGACCACTACTGCTGAATGCTGTA 3259
367 PheGlnTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGln 384
|||||
3260 CCTAGGCTACAGAGCTCTCTGTAAGAGGCCCAATTTGGACAGAGAGG 3309
384 aLThrTyrLeuGlyTyrSerLeuArgGlyGlnArgTyrPheThrGlu 400
|||||
3310 TACACTACTGGGGTACAGTTCGCGGAGCGGACGACCATGGCTGACGAG 3359
401 AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysG 417
|||||
3360 GCACGGAAGAAACTGTAGTCAGATACCGGCCCAACACAGCAACAACA 3409
417 nValArgGlnPheLeuGlyThrThrAlaGlyPheCysArgLeuThrIleProG 434
|||||
3410 AATGAGAGATTTTGGGACAGCTGGATTTTGCAGACTGTGATCCCGG 3459
434 lYpheaLThrLeuAlaAlaProLeuTyrProLeuThrLysGlnLysGly 450
|||||
3460 GGTTCGACACTTAGACGCCCACTCTACCCGCTAACCAAGAAAGAGG 3509
451 GlyPheSerTrpAlaProGlnHisGlnLysAlaPheAspAlaIleLysLy 467
|||||
3510 GAATTCCTCTGGGCTCTGACGACCAAGAGCATTGATGCTATCAAAA 3559
467 sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProP 484
|||||
3560 GGCCCTGCTGAGCGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3609
484 heThrLeuTyrValAspGlnArgLysGlyValAlaArgLysValLeuThr 500
|||||
3610 TTACCTTTATGTGATGAGCGGTAAGGAGTGGCCGAGGAGATTTAACC 3659
501 GlnThrLeuGlyProThrArgProValAlaTyrLeuSerLysLysLe 517
|||||
3660 CAACCTTAGACCATGAGAGAACCTGTGCGCTACCTGCTCAAGAACCT 3709
517 uAspProValAlaSerGlyTyrProValLysLeuLysAlaIleAlaIle 534
|||||
3710 CGATCTGTAGCCAGTGGTGGCCCATATGCTGGAAGGCTATCGACGCTG 3759
534 aAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlnLysInsIle 550
|||||
3760 TGGCCATACGTGTACAGAGCGCTGACAAATGTGCTTGGGACAGATATA 3809
551 ThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnProProAs 567
|||||
3810 ACTGTATAGCCCCCATGATGATGAGAACATGCTTGGCAGGCCCAAGA 3859
567 PArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeu 584
|||||
3860 CCGATGAGTGCACCAACGCCCGCATGACCCACTATCAAAAGCTCTCTCTCA 3909
584 hrGlnArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600
|||||
3910 CAGAGAGGTGACCTTGGTCCACACAGCCGCTCTCAACCTGCCACTCTT 3959
601 LeuProGlnGlnThrAspGlnProValThrHisAspCysHisIleLeuLeu 617
|||||
3960 CTGCTGAAAGACTGATGAACAGTACATCAGATGATGATGATCACTAT 4009
617 uIleGlnLysThrGlyValArgLysAspLeuThrAspIleProLeuThr 634
|||||
4010 GATTGAGAGACTGGGGTCCGAAAGACCTTACAGACATACCGCTAGCTG 4059
634 lYglnValLeuThrTrpPheThrAspLysSerTyrValValGlnGly 650
|||||
4060 GAGAGTGTCAACCTGCTTCACTGACGAGACACTATGTGGTGAAGGT 4109
651 LysArgMetAlaGlyAlaValValAspGlyThrArgThrIleTrpAl 667
|||||

```

```

4110 AAGAGATGCTGGGCGCGGTGTGTGAGCGGAGCCGACGATCTGGGC 4159
667 aSerSerLeuProGluGlyThrSerAlaGlnValAlaGluMetAla 684
4160 CAGACGCTCGCGAGAGAACTTCAGCACAAAAGCTGACTCATGCGCC 4209
684 euThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
4210 TCACGCAAGCTTTCGCGTCCGACAGGAATTCATTAACATTATACG 4259
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyr 717
4260 GACACGAGGTATGCTTTCGACCTGACACAGCTACATGCGCCATCTATTA 4309
717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGlu 734
4310 ACAAAGGGGTTGCTTACCTCAGCAGGAGGAAATTAAGAAACAAGAGG 4359
734 IuileuSerLeuLeuGlnAlaLeuHisLeuProLysArgLeuAlaIle 750
4360 AAATTCTAGGCTTATAGAACCGCTACATTACCAAAAAGGCTAGCTATT 4409
751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgLys 767
4410 ATACACTGCTCTGACATCAAGAAAGCTAAAGATCTCATATCCAGAGAAA 4459
767 nGluMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeu 784
4460 CCAGATGCTGACCGGTTGCCAAGCAGGAGCCGAGGTTGTAACTTC 4509
784 euProIleIleGluThrProLysAlaProGluProArgArgGlnTyrThr 800
4510 TGCCATATATAGAAATGCCCAAGCCCAAGAACCCAGACGACATACACC 4559
801 LeuGluAspTrrPngInGluIleLysLysIleAspGlnPheSerGluThr 817
4560 CTAGAGAGCTGCGAAGAGATAAAAAGATAGACCAAGTTCTGTGAGACTCC 4609
817 oGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysG 834
4610 GGAGGAGACTGCTATACCTCAGATGGGAGGAAATCTGCCCCACAAG 4659
834 IuGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThr 850
4660 AAGGTTAGAAATATGTCACAGATACATGCTTAACCCACCTGAGAACT 4709
851 LysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLe 867
4710 AAACACCTGCGACAGTTGTGTCAGAAATCCCTTATCATGTTCTGAGAGCT 4759
867 uProGlyValAlaAspSerValValLysHisCysValProCysGlnLeu 884
4760 ACCAGAGTGGCTGACTGCTGCTCAAAACATTTGTGCTGCCACAGCTGG 4809
884 aLAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySer 900
4810 TTAATGCTAATCTCCAGAAATGCTCCAGGAGGAGACCTAAGGGGAGC 4859
901 HisProGlyAlaHisThrPngInuValAspPheThrGluValLysProAla 917
4860 CACCCAGGCGCTCAGTGGAGAGTGACTCTACTAGGTAAGGCGGCTAA 4909
917 sTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrrP 934
4910 ATACGGAACAATATCATTTGGTTTGTGACACCTTTTCAGAGATGGG 4959
934 aGluAlaTyrProThrLysLysGluThrSerThrValAlaLysLys 950
4960 TAGAGGCTTATCTACTAAGAAAGAGACTTCACCGTGGTGGCTAAAAA 5009
951 IleLeuGluGluIlePheProArgPheGlyIleProLysValIleGly 967
5010 ATACTGGAAGAAATTTTCCAGATTTGGAATACCTTAAGTAAATAGGCTC 5059

```

```

967 rasPasngIyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysI 984
5060 AGACATATGATCCAGCTTTTGTGCCAGGTAAGTACGAGGAGCTGGCCAAA 5109
984 leLeuGlyIleAspTrrPlyLysLeuHisCysAlaTyrArgProGlnSer 1000
5110 TATTGGGAGATGATGTAAGAACTGCATTTGTACAGACCCCAAGCTCA 5159
1001 GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLe 1017
5160 GCACAGGTAGAGAGATGAATAGAACCATTAAGAGAACCTTACTTAATT 5209
1017 uThrThrGluThrGlyIleAsnAspTrrPheAlaLeuLeuProPheVal 1034
5210 GACCGGAGAGACTGGCGTTAATGATGTGATGATGTCCTCGCCCTTGTGC 5259
1034 euPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlu 1050
5260 TTTTATGAGGTTAGGAACACCCCTGACAGTTGGGCTGACACCCCTATGA 5309
1051 LeuLeuTyrGlyIleProProPheLeuAlaGluIleAlaPheAlaHis 1067
5310 TTACTCAGCGGGGAGACCCCCCATTTGTAAGAAATGCTTCTGTACATAG 5359
1067 rAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1084
5360 TGCTGACGTGCTGCTTCCAGGCTTGTCTCTAGCTCAAGGCACTTG 5409
1084 IuTrrPValArgGlnArgAlaTrrPlyGlnLeuArgGluAlaTyrSer... 1099
5410 AGTGGGTGAGACAAACGAGCGTGGAGCAACTCCGAGGAGGCTCACTAGGA 5459
1100 GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTy 1116
5460 GGAGGAGACTGCGAGATCCACATGCTTTTCCAGTGGAGATTCAGTCTTA 5509
1116 rValArgArgHisArgAlaGlyAsnLeuGluThrArgTrrPlyGlyPro 1133
5510 CGTTAGACGCGACCGTGCAGAGAAACCTCGAGATCGGTGGAAGGCGCTT 5559
1133 yrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIlePro 1149
5560 ATCTCGTACTTTTGACCAACCAACGCGTGTGAAGTCAAGAGAACTGCC 5609
1150 ThrTrrPLeuHisAsnHisValLysProAlaProProAspSerGln 1166
5610 ACCTGATCCATCATCCACAGTTAAACCGGCGCACCTCCGATTCGGG 5659
1166 yTrParGAlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgL 1183
5660 GTGGAAGCGCGAAGAGCTGAATAATCCCTTAAGCTTCGCTCCATCGCG 5709
1183 euValProTyrSerAsnAsnAsn 1190
5710 TGTGCTCTTACTCTGTCAATAAAC 5732

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-09-766-528-1

seq_documentation_block:

; Sequence 1, Application US/08766528

; Patent No. 6190861

; GENERAL INFORMATION:

; APPLICANT: Jay A. Fishman

; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

; NUMBER OF SEQUENCES: 74

; CORRESPONDENCE ADDRESS:

; ADDRESS: LAHIVE & COCKFIELD, LLP

; STREET: 60 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

ZIP: 02109-1875
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/766,528
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/572,645
 FILING DATE: 14-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Louis Myers
 REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: MGP-038CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)227-5941
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8060 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-766-528-1

alignment_scores:
 Quality: 5637.50 Length: 1127
 Ratio: 5.061 Gaps: 3
 Percent Similarity: 98.846 Percent Identity: 96.628

alignment_block:
 US-09-171-553B-5 x US-08-766-528-1 ...

Align seg 1/1 to: US-08-766-528-1 from: 1 to: 8060

1 GlyArgGlySerAspProLeuProGluProArgValThrLeuVal 17
 4687 GGGAGAGCGGGTTCGGACCCCTCCCGAGCCCGGGGTAACTTTGAAGGT 4736
 17 GluGlyGlnProValGluPheLeuValAspThrGlyAlaGlnHisSerV 34
 4737 GGAAGGCGCAACAGTGTGCTGCTGTTGATACCGGAGCGCAACATTGAG 4786
 34 AlaLeuLeuGlnProLeuGlyLysLeuLysGlnLysLysSerThrValMet 50
 4787 TGCTACTACAGCCATAGGAAATAAAGATAAATAATCCTGGGTGATG 4836
 51 GlyAlaThrGlyGlnArgGlnIleProThrProThrArgThrValAs 67
 4837 GGTGC.ACAGGGCAACAACAGTATCCATGACTACCCGAAG.ACAGTTGA 4884
 67 PheGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
 4885 CTGGAGATGGGACGGGTAACCCACCTGTTCTGTGTCATACCTGAGAGCC 4934
 84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
 4935 CAGACCCCTCTTAVGTAGAGACTTATTGACCAAGATGGGAGCAACAATT 4984
 101 SerPheGluGlnArgProGluValSerValAsnAsnLysProIleTh 117
 4985 TCTTTTAAACAAGGAAACAGAGTGTGCAATATACAAACCTATCAC 5034
 117 rValLeuThrLeuGlnLeuAspAspGluIleArgLeuThrSerProGln 134
 5035 TGTGTGACCTCCATTTAGATGAGATATGACATCTACTCTCCCTAG 5084
 134 AllysProAspGlnAspIleGlnSerThrLeuGlnIlePheProGlnAla 150

5085 TAAAGCTGATCAAAATATACAAATCTGTGGAGACAGTTTCCCAAGCC 5134
 151 TrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVal 167
 5135 TGGGCGAAGAACCCGAGGAGTGGGTTGGCAAGCAAGTCCCAACAGT 5184
 167 IleGlnLeuLysAlaSerAlaThrProValSerValArgGlnIleProL 184
 5185 TATCAACTGAAGCCAGTCCACACACAGTGCATGACAGATACCCCT 5234
 184 euserArgGluAlaArgGluGlyIleThrProHisValGlnArgLeuIle 200
 5235 TGAATGAAGGCTCAAGAGGAATTCGGCGCATGTCCAAAGATTAAATC 5284
 201 GlnGlnGlyIleLeuValProValGlnSerProThrAsnThrProLeu 217
 5285 CAACAGGCGCATCTAGTCTGCTGCATCTCCCTGGAATATCTCCCTGCT 5334
 217 uproValArgLysProGlyThrAsnAspThrArgProValGlnAspLeu 234
 5335 ACCGGTTAGAAAGCTGGGAGCTAATGACTATGACACAGTACAGGACTTGA 5384
 234 rglValAlaAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
 5385 GAGAGGTCAATTAACGGGTGCAGAGATATACCCCAACAGTCCCAACCT 5434
 251 TyrAsnLeuLeuSerAlaLeuProProGluIleArgAsnThrThrVal 267
 5435 TATTAACCTTGTGTGCTCTCCACCCCAACGAGGCTGATATACGATTT 5484
 267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
 5485 GGACTTAAGAGATGCTTCTGCTGCTGATGATACACCCACAGCCCAAC 5534
 284 roLeuPheAlaPheGluIleThrArgAspProGlyThrGlyArgThrGlyGln 300
 5535 CACTTTTTCCTTCGAATGAGATCCAGATCCAGGAGAAACCGGCGAG 5584
 301 LeuThrTriThrArgLeuProGlnGlyPheLysAsnSerProThrIleP 317
 5585 CTACCTGACCCGAGCTGCTCCCAAGGCTTCAGAACTCCCAACCTCTT 5634
 317 eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
 5635 TGACGAAAGCCCTACACAGACCTGGCCCAACTTCAGAGTCAACACCTC 5684
 334 lValThrLeuLeuGlnIleThrValAspAspLeuLeuAlaGlyAlaThr 350
 5685 AGGTGACCTCTCCAGTACGTGAGTACCTCTTGTGGGAGGACAC 5734
 351 LysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuSerAs 367
 5735 AAACAGAGCTGCTTGAAGAGCCAGAGGCACTACTGCTGGAATGTCTGA 5784
 367 PheGlyIleThrAlaSerAlaLysLysAlaGlnIleCysArgArgGlu 384
 5785 CCAAGGCTACAGAGCTCTGTCAAGAGCCCAAGTTTGACAGAGAGAGG 5834
 384 AlThrThrLeuGlnIleThrSerLeuArgGlyGlnArgThrLeuGlu 400
 5835 TAACATCTTGGGTACAGATTACGGGAGCGGACGAGTGTGACGAG 5884
 401 AlaArgLysLysThrValAlaGlnIleProAlaProThrThrAlaLysG 417
 5885 GCACGGAAGAAACGTAGTACAGATACCGGCGCCCAACACCAACCAACA 5934
 417 nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuThrIleProG 434
 5935 AATGAGAGAGTTTGGGAGACGCTGAGANTTTGACAGCTGTGATCCCGG 5984
 434 lPheAlaThrLeuAlaAlaProLeuIleProLeuThrLysGlnLysGly 450
 5985 GGTTCGACCTTAGACAGCCCACTACCCCTAACCAAGAAAGAGG 6034

451 GlyPheSerTrpAlaProGlnHisGlnLysAlaPheAspAlaIleLysLys 467
|||||
6035 GAATTCCTCGGCTCCTGAGCACCAGAAAGCATTTGATGCTATCCAAAAA 6084
467 sAlaLeuLeuSerAlaProAlaIleAlaLeuProAspValThrLysPro 484
|||||
6085 GGCCTGCTGAGCGCAGCTGCTGGCCCTCCCTGACCTAGCTAAACCTT 6134
484 heThrLeuTrpValAspGluArgLysGlyValAlaArgGlyValLeuThr 500
|||||
6135 TTACCTTTATGTGATGAGCGTAAGGAGTAGGCCGGGAGTTTAAAC 6184
501 GlnThrLeuGlyProTrpArgArgProValAlaArgLysSerLysLys 517
|||||
6185 CAACACCTTAGAGACCATGAGAAAGACCTGCTCCCTACCTGTCATAAGAACT 6234
517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaLay 534
|||||
6235 CGATCCTGTAGCCAGTGGTGGCCCATATGCTGTAAGGCTATCCGAGCTG 6284
534 aAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle 550
|||||
6285 TGGCCATCTGCTGCAAGGACGCTGACAAATTGACTTTGGACAAAGATAT 6334
550 eThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnProPro 567
|||||
6335 AACTGTATAGCCCCCATGCAATTGGAGAACATCTTGCGGACGCCCCAG 6384
567 sPArgTrpMetThrAsnAlaArgMetThrHisTrpGlnSerLeuLeu 583
|||||
6385 ACCGATGATGACCAACGCCCATGACCCATCAAGCCTGCTTCTC 6434
584 ThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLe 600
|||||
6435 ACAGAGAGGGTCACGTCGCTCCACCAACCGCTCAACCTGCGACACT 6484
600 uLeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeu 617
|||||
6485 TCTGCTGAGAGACTGATGAACAGTGTCTATGATGCTCAACTCAACTAT 6534
617 euIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThr 633
|||||
6535 TGATTGAGAGAGCTGGGGTCCGCAAGACCTTACAGACATACCGCTGACT 6584
634 GlyGluValLeuThrTrpPheThrAspGlySerSerTrpValValGluG 650
|||||
6585 GGAGAGAGTGTAACTGCTTCACTGACGAGAACAGCTATGTGTGGAAGG 6634
650 yLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgThrIleTrp 667
|||||
6635 TAAAGAGATGGCTGGGGCGGGCGGTGTGAGCGGACCCGACGATCTGGG 6684
667 laSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAla 683
|||||
6685 CCAGCAGCTGCGCGGAGGAACTTCAAGCAAAAGGCTGAGCTCATGGCC 6734
684 LeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleThr 700
|||||
6735 CTCACCAAGCTTTCGGCTGGCGGAGGAAATCCATTAACATTTATAC 6784
700 rAspSerArgTrpAlaPheAlaThrAlaHisValHisGlyAlaIleTrp 717
|||||
6785 GGCACAGAGGTATGCTTGGCAGCTGCACACAGTACATGGGGCATCTATA 6834
717 ySlnArgGlyLeuLeuThrSerAlaGlyArgGluLysLysGlu 733
|||||
6835 AACAAAGGGGGTGTCTTACCTCAGCGAGGAGGGAATAAAGCAAAAGAG 6884
734 GluIleLeuSerLeuGluAlaLeuHisLeuProLysArgLeuAlaIle 750
|||||
6885 GAAATTCCTAAGCCTATTAGAACCGGTACATTTTACCAAAAAGGCTACTAT 6934

750 eIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGly 767
|||||
6935 TATACACTGCTCTGGACATCAGAAAGCTTAAAGTCTCATATCCAGAGAA 6984
767 sngIleMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeu 783
|||||
6985 ACCAGATGGCTGACCGGTTGGCCAGACAGACCCGAGGTTTAACTT 7034
784 LeuProIleLeuGluThrProLysAlaProGluProArgArgLysThr 800
|||||
7035 CTGCTATATATAGAAATATGCCAAAGCCCAAGAACCCAGACGACATAC 7084
800 rLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThr 817
|||||
7085 CTTAGAAAGATGGCAAGAGATTAAGAAAGATAGACCA.TTCTGTGAGACATC 7133
817 roGluGlyThrCysTrpThrSerTrpGlyLysGluIleLeuProHisLys 833
|||||
7134 CGGAGGAGACCTGCTATACCTCAGATGGGAAAGAAATCCGCCACAA 7183
834 GluGlyLeuGluTrpValGlnGlnIleHisArgLeuThrHisLeuGly 850
|||||
7184 GAAGGTTAGATATATGTCACAAGATACATCTGCTACCCACCTAGGA 7233
850 hTrpHisLeuGlnGlnLeuValArgThrSerProTrpHisValLeuArg 866
|||||
7234 CTAACACCTGCAAGCACTTGTGAGAACATCCCTTATCATGTCTGAGG 7283
867 LeuProGlyValAlaAspSerValValLysHisCysValProCysGln 883
|||||
7284 CTACACGAGAGTGGCTGACCTGGTGTCAAAACATTTGTGCTGCCACGT 7333
883 uValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgLys 900
|||||
7334 GGTAAATGCTATCTCTCCAGAAATGCTCCAGGAAAGACCTAAGGGAA 7383
900 eHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAla 916
|||||
7384 GCCACCGAGCGGCTCACTGGGAAAGTGAATCTTCACTGAGTAAGCCGCT 7433
917 LysTrpGlyAsnLysTrpLeuLeuValPheValAspThrPheSerGlyTr 933
|||||
7434 AAATATGGAACAATATCTATGTTGTTTGTAGACACCTTTTCAGAGATG 7483
933 pValGluAlaTrpProThrLysLysGluThrSerThrValValAlaLysL 950
|||||
7484 GGTAGAGGCTTATCTACTAGAAAGAACTTCAACCGTGTAGCTAAAA 7533
950 ySlnLeuGluGluIlePheProArgPheGlyIleProLysValIleGly 966
|||||
7534 AAATACTGGAAGAAATTTTCCAAAGATTTGGAATACCTTAAGTAATAGG 7583
967 SerAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 983
|||||
7584 TCAGCAATGTGTCACGCTTTGTTGCCAGGTAAGTCAGGGACCGGCCAA 7633
983 sIleLeuGlyIleAspTrpLysLeuHisCysAlaTrpArgProGlnSer 1000
|||||
7634 GATATTGGGATGTGATTGGAACCTGCAATGTGCATACAGACCCCAAGCT 7683
1000 eArgGlyValGluArgMetAsnArgThrIleLysGluThrLeuThrLys 1016
|||||
7684 CAGCAGAGTAGAGAGATGATGAAACCAATTAAAGAACCTTACTAATA 7733
7734 TTGACCGGGAGAGACTGGGTTATGATGATGATGATGATGATGATGATG 7783
1017 LeuThrTrpGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
|||||
1033 lLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTrp 1050
|||||
7784 GCTTTTATGAGTTAGAACACCCCTGACAGATTTGGGTGACCCCTATG 7833
1050 lLeuLeuTrpGlyLysProProLeuAlaGluIleAlaPheAlaHis 1066


```
|||||
7834 AATTAAGTCTACGGGGGACCCCCCATGTTGTAAGAAATGCTTGTACAT 7883
1067 SerAlaAspValIleuLeuSerGlnProLeuPheSerArgLeuLysAla 1083
7884 AGTGTGATGTCTGCTTCCAGCCTTGTCTCTAGGCTCAAGCACT 7933
1083 uGluTrpValAlaArgIleAlaTTPlysGlnLeuArgLysAlaTyrSer 1099
7934 TGAGTGGGTGAGACACAGCCTGGAGGCACTCCGGAGGCTTACTGAG 7983
1100 ..GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerVal 1115
7984 GAGGAGGAGACTTGCAGATCCCATGCTTCCCAAGTGGAGATTCACTG 8033
1116 TyrValAlaArgHisArgAlaGlyAsn 1124
8034 TACGTTAGCGCCACCGTGCAGGAAC 8060

seq_name: /cgn2_6/prodata/2/ina/bh_COMB.seq:us-08-766-528-2

seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MCP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-2

alignment_scores:
Quality: 4489.50 Length: 1191
Ratio: 5.033 Gaps: 2
Percent Similarity: 74.895 Percent Identity: 74.643

alignment_block:
US-09-171-553b-5 x US-08-766-528-2
Align seg 1/1 to: US-08-766-528-2 from: 1 to: 7333
```

```
1 GlyArgArgLysSerAspProLeuProGluProArgValThrLeuLysVal 17
2173 GGGAGAGGGGGTTGGACCCCCCTCCGAGACCCAGGTAACCTTTGAAAGT 2222
17 GlnGlyGlnProValGluIleLeuValAspThrGlnGlyAlaGlnLysSerVal 34
2223 GAGAGGGGCAACCAAGTTGATGTTCTGGTTGATACCGGAGCGGACATTCAG 2272
34 alleuLeuGlnProLeuGlnLysLeuLysGlnLysLysSerTrpValMet 50
2273 TGTGCTTACCAACCAATTAGGAAACCTAAAGAAAAAAATCTGGGTGAGG 2322
51 GlyAlaThrGlnGlnArgGlnTyrProTyrThrArgArgTrpValAs 67
2223 GGTGCCACAGGCGACAGCGGAGTTCATGACTACCCGAGAGACCGTTGA 2372
67 pleuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
2373 CTGGGAGGTGGAGCGGTTAACCCACTCGTTTGTGTCATCCCTGAGTGGC 2422
84 roValProLeuLeuGlnArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
2423 CAGTACCCCTTAGGTAGAGACTTACTGACCAAGATGGAGCTCAAAAT 2472
101 SerPheGlnGlnGlnArgProGluValSerValAsnLysProIleTh 117
2473 TCTTTGAAACAAGAGACAGACAGAGTCTGTGAATAAACCAACCTAC 2522
117 rValLeuThrLeuGlnLeuAspGlnLysArgLeuTyrSerProGlnV 134
2523 TGTGTTGACCCCTCCAAATGATGATGAATATGACATATATCTCCCAAG 2572
134 allProAspGlnAspIleGlnSerTrpLeuGlnGlnPheProGlnAla 150
2573 TAAAGCTGATCAAGATATACAGTCTGTTGGAGACGTTTCCCAAGCC 2622
151 TrpAlaGlnThrAlaGlnLysMetGlyLeuAlaLysGlnValProGlnV 167
2623 TGGGCGAAGAACCGCAGGAGTGGGTTTGGCAAGCAAGTTCCCCACAGT 2672
167 IleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrProL 184
2673 TATTCAACTGAAGGCCAGTGTACACACAGTATCACTGACAGATTAATC 2722
184 euSerArgGlnAlaArgGlnGlyIleTrpProHisValGlnArgLeuIle 200
2723 TGAGTAGAGAGGCTCGAGAGGAATTTGGCCGATGTTCAAGATTAATC 2772
201 GlnGlnGlnIleLeuValProValGlnSerProTrpAsnThrProLeu 217
2773 CAACAGGCGATCCTAGTCTGCTCCCAATCCCTTGGAATACCTCCCTCT 2822
217 uProValArgLysProGlnThrAsnAspTyrArgProValGlnAspLeu 234
2823 ACCGGTTAGGAAGCCTGGAGCAATGATTAATCGACCACTACAGGACTGA 2872
234 rGlnValAsnLysArgValGlnAspIleHisProThrValProAspPro 250
2873 GAGAGGTCATTAAGAGGTGCAGAGCATACACCCACCGTCCGAAACCT 2922
251 TyrAsnLeuLeuSerAlaLeuProProGlnArgAsnTrpTyrThrVal 267
2923 TATAACCTCTTGAGCGCCCTCCGCTGACGAACTGTTACACAGATAT 2972
267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
2973 GGACTTAAAGATGGCTTCTTGTGCTGAGATTACACCCCACTAGCCAC 3022
284 roLeuPheAlaPheGlnTrpArgAspProGlnTyrGlnArgThrGlyGln 300
3023 CACTTTTACTCTGGAATGAGAGATCCAGGTACGGAGAGAACCGGGCAG 3072
```

```

301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePhe 317
|||||
3073 CTCACCTGGACCCGACTGCCCCAGAGGTTTCAAGACTCCCCGCCATCTT 3122
|||||
317 eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
|||||
3123 TGAAGAAAGCCCTACAGAGGAGCTGGGCCAATTCAGATCCAAACACCCTC 3172
|||||
334 LysValThrLeuLeuGlnThrValAspAspLeuLeuAlaGlyAlaThr 350
|||||
3173 AGGTGACCTCTCCACGTACGTGATGACCTGCTTGGGGGAGGACACC 3222
|||||
351 LysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuGlnLeuSerAs 367
|||||
3223 AACAGGACTGCTTAGAAGGTACGAAAGCCACTGCTGGAATTGCTGTA 3272
|||||
367 PLeuGlyThrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluV 384
|||||
3273 CCTAGGCTACAGAGCTCTGCTAAGAGGCCAGATTTCAGAGAGAGAGG 3322
|||||
384 aLThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGlu 400
|||||
3232 TTAACATCTTGGGGTACAGTTTGGGGGCGGCGACGATGCTGACGCGAG 3372
|||||
401 AlaArgLysLysThrValIleGlnIleProAlaProThrThrAlaLysG 417
|||||
3373 GCACGGAAAGAAACTGTAGTCCAGATACGGCCCCAACCCACAGCAACACA 3422
|||||
417 nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProG 434
|||||
3423 AGTGAGAGAGTTTGGGGACAGCTGGATTTGGCAGACTGTGATCCCGG 3472
|||||
434 LyrPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlnLysGly 450
|||||
3473 GGTGTGGACCTTAGCAGCCCACTTAACCGGCTAACCCAAAGAAAGGG 3522
|||||
451 GlyPheSerTrpAlaProGlnHisGlnLysAlaPheAspAlaIleLys 467
|||||
3523 GGTGG..... 3527
|||||
467 sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysPro 484
|||||
3527 ..... 3527
484 nethrLeuTyrValAspGluArgLysGlyValAlaIleArgGlyValLeuThr 500
|||||
3527 ..... 3527
501 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLys 517
|||||
3527 ..... 3527
517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaIle 534
|||||
3527 ..... 3527
534 aLAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlnAsnIle 550
|||||
3527 ..... 3527
551 ThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAs 567
|||||
3527 ..... 3527
567 PArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuT 584
|||||
3527 ..... 3527
584 hArgLysValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600
|||||
3527 ..... 3527
601 LeuProGlnGlyThrAspGluProValThrHisAspCysHisGlnLeuLe 617
|||||
3527 ..... 3527
617 uLleGlnGlyuThrGlyValArgLysAspLeuThrAspIleProLeuThrG 634
|||||
3527 ..... 3527
634 LysLysValLeuThrTrpPheThrAspGlySerSerTyrValValGlnGly 650
|||||
3527 ..... 3527
651 LysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgThrIleTrpAl 667
|||||
3527 ..... 3527
667 aSerSerLeuProGlnGlyThrSerAlaGlnLysAlaGlnLeuMetAlaL 684
|||||
3527 ..... 3527
684 euThrGlnAlaLeuArgLeuAlaGlnGlyLysSerIleAsnIleTyrThr 700
|||||
3527 ..... 3527
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrI 717
|||||
3527 ..... 3527
717 sGlnArgGlyLeuLeuThrSerAlaGlyValArgGlyLysAsnLysGlnG 734
|||||
3528 ..... 3527
3528 ..... 3527
734 LysLeuSerLeuLeuGlnAlaLeuHisLeuProLysArgLeuAlaIle 750
|||||
3565 AAATTCTAAGCCATATAGAAGCCTTACATTGTCCAAAAAGCCTAGCTATT 3614
|||||
751 LLeHisCysProGlnHisGlnLysAlaLysAspLeuIleSerArgGlyLys 767
|||||
3615 ATCACTGTCTGTCGACATCAGAAAGCCAAAGATCTCATATGTCAGAGGAA 3664
|||||
767 nGlnMetAlaAspArgValAlaLysGlnAlaIleAlaIleValAsnLeu 784
|||||
3665 CCAGATGGCTGACCGGGTGGCCAGCAGCAGCCCGGCTGTAACTTCC 3714
|||||
784 euProIleIleGlnThrProLysAlaProGluProArgArgGlnTyrThr 800
|||||
3715 TGCCCTAATATAGAAGCCCAAGCCCAAGCCCAAGCCCAAGCAGTACACC 3764
|||||
801 LeuGluAspTrpGlnGlnLysLysIleAspGlnPheSerGlyuThrPr 817
|||||
3765 CTGAAGAGACTGGCAGAGATAAAGAAAGATAGACAGATCTCTGAGACTCC 3814
|||||
817 OGlnGlyThrCysTyrThrSerTyrGlyLysGlnLysLeuProHisLysG 834
|||||
3815 GGAAGGGAGCTGCTATACCTCATATGGAAGAAATCTGCCCAACAAAG 3864
|||||
834 LysGlyLeuGlnTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThr 850
|||||
3865 AAGGGTAGAATATGTCCAAAGATACATCGTATACCCACCTAGGAAGT 3914
|||||
851 LysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLe 867
|||||
3915 AACACCTGCGAGCAGTGGTCAGAACATCCCTATACATGTTCTGAGGCT 3964
|||||
867 uProGlyValAlaAspSerValAlaLysHisCysValProCysGlnLeuV 884
|||||
3965 ACCAGAGTGGCTGACTGGTGGTCAAACTTGTGGCTTGCACACTGG 4014
|||||
884 aLAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySer 900
|||||
4015 TTATATGCTATCTTCCAGAAATACCTCCAGAAAGAGACTAAGGGGAGAC 4064
|||||
901 HisProGlyValAlaHisTrpGlnValAspPheThrGlnValLysProAla 917
|||||

```

4065 CACCCAGCGCTCACTGGGAAGTGAAGTCACTGAGGTAAAGCCGGCTAA 4114
 917 STYGLVSNLYSTYRLEULEUVALPHEVALSPHTRPHESEGLYTRPV 934
 4115 ATAGGAAACAATATCATGTTGTTTGTAGACACCTTTTCAGATGGG 4164
 934 ALGUALATYRPROTHRLYSGLYUTHRSETRVALVALALYSLS 950
 4165 TAGAGGCTTACTACTAAAAAGACCTCAACCGTGGGTAAAGAA 4214
 951 ILEUUGLULIPEPROAIRPHEGLYILEPROLYSVALILEGLYSE 967
 4215 ATACTGGAGAAATTTTCCAGATTGGATTACCTAAGTAATAGGCTC 4264
 967 TASPASNGLYPROALAPHEVALAGLVALISERGLYLEUALAYS 984
 4265 AGACAATGGTCCAGCTTTCGTTGCCAGTAAGTCAAGGACTGGCCAGA 4314
 984 ILEUUGLYLEASPTPLYSLEUHSYSAATYRATPROGLINSE 1000
 4315 TATGGGGATTGATTGAAACTGCATTGCTATACAGAACCCCAAGCTCA 4364
 1001 GLYGLVALGLUARGMETSNARGTTHRLYSGLUTHRLEUHLYSLE 1017
 4365 GGACAGGTAGAGAGATGATAGACCATTAAGAGACCTTACCAATAT 4414
 1017 UTHRTHGLUTHRGLYILEASNAAPTPEMTALALEULEUPROPH 1034
 4415 GACCACAGAGACTGGCATTAATGATTGATGGCTCTCTCCCTTGTGC 4464
 1034 EUPHEARGVALARGASNTHRPROGLYINPHEGLYUTHRPROTYR 1050
 4465 TTTTAAAGGTGAGAACACCCCTGGACAGTTGGGCTGACCCCTATAA 4514
 1051 LEULEUTYRGLYGLYPROPROPHEUVALAGLULI EALPHEALHISSE 1067
 4515 TTGGCTACGGGGGACCCCGCCGTTGGCGAATATGCTTTCACATAG 4564
 1067 PALLASPVALLLEUUSERGLINPROLEUPHESERARGLEUVAL 1084
 4565 TGCTGATGCTGCTCTTCCAGCCTTGTCTCTAGGCTCAAGGCCCTGC 4614
 1084 LUTRPVALARGVALARGALATRPYSLINLEUARGLULIARYSER 1100
 4615 AGTGGGTGAGGACGACGAGCTGAGAGAGCTCCGGAGGCTTACTAGA 4664
 1101 GLYASPLEUGLINALPROHISARGPHEGLINVALGLYASPER 1117
 4665 GGAGACTTGGCAAGTCCACATCGCTTCCAAAGTTGGAGATTCACTAT 4714
 1117 LARGARGHISARGALAGLYASLEUGLUTHRARGTRPLYSGLY 1134
 4715 TAGACGGCCACCGTGCAGAAACCTCGAGACTCGTAGAAGGACCTTAT 4764
 1134 EUVALLEULEUTHRTHRPROTHRALVALYSVALGLUGLYLEPRO 1150
 4765 TCGTACTTTGACACACCAACGCGTGTGAAATCGAAGG 4804
 1151 TRPILHISALASERHISVALYSVALPROALPROPROASPSE 1167
 4804 4804
 1167 PARGALAGLULYTHRGLUASNPROLEUYSLEUATGLUHSARGLE 1184
 4805 4805
 1184 ALPROTYRSEASNAASNAUSER 1191
 4836 TTTCTTACTCTGTCAATACTCC 4858
 seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-075-272-1
 seq_documentation_block:

Sequence 1, Application US/09075272
 Patent No. 6136598
 GENERAL INFORMATION:
 APPLICANT: MILLER, A. DUSTY
 APPLICANT: WOLGAMOT, GREG
 APPLICANT: BONHAM, LYNN
 TITLE OF INVENTION: MDS DUNNI ENDOGENOUS RETROVIRAL
 TITLE OF INVENTION: PACKAGING CELL LINES
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: U.S.A.
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/075,272
 FILING DATE: 08-MAY-1998
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/046,140
 FILING DATE: 09-MAY-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: POOF, Brian W.
 REGISTRATION NUMBER: 32,928
 REFERENCE/DOCKET NUMBER: 14538A-003710
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 467-9600
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8655 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-09-075-272-1
 Alignment_scores:
 Quality: 4339.50 Length: 1202
 Ratio: 4.201 Gaps: 8
 Percent Similarity: 85.940 Percent Identity: 68.968
 Alignment_block:
 US-09-171-553B-5 x US-09-075-272-1 ..
 Align seg 1/1 to: US-09-075-272-1 from: 1 to: 8655
 1 GYATGARGGLYSERASPPROLEUPROGLUPROARGVALTHRL 17
 2291 GGAAAGTCGGGCTCAGACCCCTCCCTGAGCCTAGGTAACCTTGTGCTG 2340
 17 LGUGLYGLNPROVALGUPHELEUVALASPTTHGLVALAGL 34
 2341 GGAGGGAGCTCCGTTAATTTCTGATGATGAGAGCAGAGCATTCGG 2390
 34 ALLEULEUGLNPROLEUGLYLSLEUYSGLULYSLSSETRPV 50
 2391 TACTTACCAAGCCCTGGGCAACTAGGCTCTAAAGAGCATTAGTGTT 2440
 51 GYALATHRGLYGLNARGGLNTRPOTRPTHRTHRARGTRVAL 67
 2441 GGAGCACTGGAGTAACCTTACCTCGACACCAAAAGAGCCTTACA 2490
 67 PLEUGLYVALGLYARGVALTHRHISERPHLEUVALILEPROGL 84


```

684 eUThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
|||||
4335 TTACAGAGCCCTCCGAGAGGCCGAGATGATCATATATTTTACACT 4384
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrThr 717
|||||
4385 GACAGCGCCCTATGCTTTCTTCTACTGACATATCATGGGGCCATCTACAG 4434
717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGlu 734
|||||
4435 GCAGCGAGGGTTATTGACTTTCAGCCGGTAAAGACATCAAAACAAGAAG 4484
734 IuileuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIle 750
|||||
4485 AAATTTCTGGCCCTGTAGAGCCATACGCCCAAGAGGTACCATC 4534
751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAs 767
|||||
4535 ATCCACTGTCCTCCGTCACCAAGAGAGGACTTATGCGCCAGAGGCA 4584
767 nGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeu 784
|||||
4585 CCGGATGCGAGACTCAGTACAAACAGGTTCTCAGGGGGCCATGATCT 4634
784 euProIleIleGluThrPro...LysAlaProGluProArgArgGlnTyr 799
|||||
4635 TAACGTAAAGCAAAATCCGTCACAAAGCCCTGAG..... 4669
800 ThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluTh 816
|||||
4670 .....GATGAACATGATGATATAAAGAACATATTTGGAC 4704
816 TrpGluGlyThrCysTyrThrSerTyrGlyLysGluIleuProHisL 833
|||||
4705 TAGTATCCCTCCCATCTTTTCGAGGAAATAAGACTTGACACCCG 4754
833 ySglnGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 849
|||||
4755 AGGAAGAAATTAATTTGTAAGAGGATACACCAAGTTTACCCACTGGA 4804
850 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuAr 866
|||||
4805 GTTGAATAAATGATGAGACTGATTAAGTAAGTCCCGTATCAAGTCCCTAA 4854
866 gluProGlyValAlaAspSerValValLysHisCysValProCysGln 883
|||||
4855 CTGGAAGTCAGTGGCTCAAAAGATTAACCTCTCAAGCGGTGTGAT 4904
883 euValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGly 899
|||||
4905 TCACTAATGCAACTAAACCTTCAAGAAAGACCTGGAAAGAGACACGGGGA 4954
900 SerHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAl 916
|||||
4955 GACCGTCTCGAGGTATTGGAGGTGATTTTACTGAAGTTAAACCCGG 5004
916 aLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyT 933
|||||
5005 AATGATGTGATTAAGTATCTGTTAGTATTGTGACACCTTTTCAGGAT 5054
933 rPValGlnAlaTyrProThrLysLysGluThrSerThrValValAlaLys 949
|||||
5055 GGGTTGAAGCGTCCCACTAAACTGAGACTGCCAGGATGGCCAG 5104
950 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleG 966
|||||
5105 AAGATTTTGAAGAATCTCCCAAGATATGAGTACTTAAGTATGCG 5154
966 ySerAspAsnGlyProAlaPheValAlaGlnValSerGlnLysLeuAla 983
|||||
5155 GTCCGACAAATGAGACACCTTGTGGCCAGTAACTAGGATTTGGCA 5204
983 ySIIleuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSer 999
|||||
5205 CTCAGTTGGGCAATGATGGAATTAATCATGTGTGCTTACCCCTCAAGC 5254
1000 SerGlnValAlaGluArgMetAsnArgThrIleLysGlnThrIleThrLy 1016
|||||
5255 TCAGGACAGGTAGAGAGATGAATAGAACATTAAGAGACCTTGATTA 5304
1016 sLeuThrThrGluThrGlyLysAsnAspTrpMetAlaLeuLeuProPheV 1033
|||||
5305 ATTGCGCATGAGACCGCGGGAAGACTGGGTGGCTCTCTCCCTCC 5354
1033 aLleuPheArgValArgAsnThrProGlyLysIlePheGlyLeuThrProTyr 1049
|||||
5355 CGCTTTCCGAGCCCGGAGACACCCCGAGGATTTGGGCTCACCTCTTT 5404
1050 GluLeuLeuThrGlyGlyLysProProProLeuAlaGlu.....IleAl 1063
|||||
5405 GAATTTCTGATGGGGGACCTCCCTTTAATAAAGATGTGGACATT 5454
1063 aPheAlaHisSerAlaAspValLeuLeuSerGlnProLeuPheSerArgL 1080
|||||
5455 GGTTCGCCGATTCAGGCTCTGCTTACCTCTCT...TTCCTTATTCATT 5501
1080 euLysAlaLeuGluTyrValArgGlnArgAlaTrpLysGlnLeuArgGlu 1096
|||||
5502 TAAAGGCCCTGAAGATTAATAGACCCCAATTTGGGACCACTTAAGACG 5551
1097 AlaTyrSerGlyGlyAspLeuGlnValProHisArgPheGlnValGlyAl 1113
|||||
5552 GCCTACACCCAGGAGGACCCGACGAGTACCCACAGAGTCCAGGTCGGGA 5601
1113 pSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpL 1130
|||||
5602 CCAAGTCTTGTCAGACGACATCGAACCGGTAGCTTGAACCAACGGTGA 5651
1130 ySglnProTyrLeuValLeuLeuThrThrThrProThrAlaValLysValGlu 1146
|||||
5652 AGGAGCCCTATTAGTGTACTAACAACCTCTCAGGACGAGTGAAGTTGAC 5701
1147 GlyIleProThrTrpIleHisAlaSerHisValLysProAlaProProP 1163
|||||
5702 GGGATTCCTCTCTGATCCAGCTTCAACGCTCAACAGGAGGCCCTCACTCA 5751
1163 oAsp.....SerGlyTrpArgAlaGluLysThrGluAsn 1175
|||||
5752 AGATGAAGAAACCCAGCAAGACAACTGGGCTGGAAGCCACTGATTAACC 5801
1175 rGluLeuAspArgLeuHisArgLeuValProTyrSerAsnAsnSer 1191
|||||
5802 CTCTTAAGCTTGGCTTGGCGAGAGACCCCTC.....CATCAC 5842
1192 ProGly 1193
|||||
5843 CTTGGA 5848
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1
seq_documentation_block:
; Sequence 1, Application US/08716351A
; Patent No. 6033905
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Glibon Ape Leukemia Virus-Based
; TITLE OF INVENTION: Retroviral Vectors
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716, 351A
; FILING DATE:

```



```

3911 CCTTTATTGAGCTGAGAACATCAGACGCTTTTGACACATATAAAAA 3960
467 salaleuLeuSerAlaProAlaLeuAlaLeuProAspValThrIysProP 484
|||||
3961 AGCGTGTGCTGACGCGCCCTGCATTTGGCCCTCCCAAGACTCACAGACCAT 4010
484 hehIleuTyRValAspGluArgLysGlyAlaAlaArgLysValLeuThr 500
|||||
4011 TCACCTATATATAGATGAGAGAGCGGCGTGGCCCGGGGAGTCTCACT 4060
501 GlnThrLeuGlyProTyrArgArgProValAlaTyrLeuSerLysLysLe 517
|||||
4061 CAGACTTTAGACCTGGCGGGCGGCGAGTACATATATCAAAAAAACT 4110
517 uAspProValAlaSerGlyTyrProValCysLeuLysAlaIleAlaIle 534
|||||
4111 GGATCCGGTGGCCAGCGGGTGGCCACCTGCTGAAAGCGTTCCACAG 4160
534 aIAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle 550
|||||
4161 TACACTCTCTCTCAAAAGACGTATTAACCTTGGGACAAATATGTG 4210
551 ThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAs 567
|||||
4211 ACTGTGATTTGCTTCCCATAGCCTGGAAGCATCTGGCGCAACCCCGCA 4260
567 PArgTyrMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeu 584
|||||
4261 CCGGTGATGACCAATGCGCAATGACTCATTTACAGAGCTGTGTATA 4310
584 hrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600
|||||
4311 ATGAAGAGGTATCTCTTTGGCCCTCTGCTGCTTAAACCCAGTACCCCA 4360
601 LeuProGlnGluThrAspGluProValThrHisAspCysHisGlnLeu 617
|||||
4361 CTTCAGTCGAGTCGGAAGCCACCCAGTCACAGGTCGCTCAGAAATTCCT 4410
617 uIleGlnGluThrGlyValArgLysAspLeuThrAspIleProLeuThrG 634
|||||
4411 CGCGGAAGAACTGGAAGTACGACGACCTGAGAACCAACCAATGTCGCCG 4460
634 LysGluValLeuThrPheThrAspGlySerSerTyrValValGlnGly 650
|||||
4461 GG...GTGCCAACCTGATATACAGACGATGACGTTTCAACAGGAAGGT 4507
651 LysArgMetAlaGluAlaAlaValAlaAspGlyThrArgThrIleTyrAl 667
|||||
4508 AAAGGAGACAGCGGCGCCGATCTAGATGCGAAGCGGAGGTATGGGC 4557
667 aSerSerLeuProGlnGlyThrSerLysGlnLysAlaGluLeuMetAlaL 684
|||||
4558 TAGCAGCGCTGCCAAGAGGTACGACCCGAGAAAGGCTGAATAGTACCT 4607
684 euThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
|||||
4608 TGACGAGGAGCATTAAGCTGGCGGAGAAAGAAACATCACTTACACG 4657
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrL 717
|||||
4658 GACAGAGGATATCTTTTGCACCTGCTCATTTCAATGGGCAATATATA 4707
717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGlnIleLysAsnLysGlnG 734
|||||
4708 GCAGAGGGGGCTGCTCACTTCTGCTGGAAGATATCAAAAAAAGAGGG 4757
734 IuIleuSerLeuLeuGlnAlaLeuHisLeuProLysArgLeuAlaIle 750
|||||
4758 AAATTTGGCCTGCTAGAGGCATTCATTCCTAGCGGGTGCCTATT 4807
751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAs 767
|||||
4808 ATCCACTGTCTGTGGCCACGAGGGGAGTAACCTGTGGCCACTGGGAA 4857

```

```

767 nGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeuL 784
|||||
4858 CCGAGAGCGCCGAGAGAGCTGCAAAAGCAAGCCCGCTGTCAGACAGATGC 4907
784 euProIleIleGluThrProLysAlaPro...GluProArgArgGlnTyr 799
|||||
4908 TGGCAGGAACTAACAACCTCAAGAGCAATTCGAGCGCGCTCAAGAAAG 4957
800 ThrLeuGluAspTyrGlnGluIleLysLysIleAspGlnPheSerGluTh 816
|||||
4958 ACCAGCGCG... 4966
816 rProGlnGluThrCysTyrThrSerTyrGlyLysGluIleLeuProHisL 833
|||||
4967 .....AGGAGCTCAACCCCTGACC 4985
833 LysGlnGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 849
|||||
4986 GG...GCAAAAGAAATTCATTAGCGGTACATCAGTTAACTCACTTAGA 5032
850 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuAr 866
|||||
5033 CCAGAAAGACTTCTTCCAACTAGTAGAACCGTACAGCCCTCTCATCCCGAA 5082
866 gluProGlyValAlaAspSerValValLysHisCysValProCysGlnL 883
|||||
5083 CCTCAATCTCACTGAGTTCGCAAGTCAACAGTCAAGTGTGAGGCTTGAGCA 5132
883 euValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGly 899
|||||
5133 TGACTAATGGCGTACACCACTACAGAGAGACCGGAAAGGCAACAGAGGA 5182
900 SerHisProGlyAlaHisTyrGluValAspPheThrGluValLysProAl 916
|||||
5183 GATGACCCGCGGTGACTGGAGGTAGACTTCAAGAAATTAAGCTCG 5232
916 aLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyT 933
|||||
5233 TCGGTATGGAACAAAGATATCTGTAGTATTCATGATCTTCTCCGAT 5282
933 rPValGlnAlaTyrProThrLysLysGluThrSerThrValValAlaLys 949
|||||
5283 GGGTAGAAGCTTTCTTACCAAAACGAAACGCGCTATATCTCTGTAAA 5332
950 LysIleLeuGlnGluIlePheProArgPheGlyIleProLysValIleG 966
|||||
5333 AAATATATTAGAAAGAAATTTACCCCTTGGGATCCCTAAGTACTCGG 5382
966 ySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaL 983
|||||
5383 GTCAGACAAATGGCGCGCTTGTGCTGCTAGGTAAGTCAAGGAGCTGGCA 5432
983 LysIleLeuGlyLysPThrLysLeuHisCysAlaTyrArgProGlnSer 999
|||||
5433 CTCAACTGGGCAATAATTGGAAGTTACATTTGGCTATAGACCCAGAGGC 5482
1000 SerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrL 1016
|||||
5483 TCAGGTCAAGTAGAAAGTAAGTAACAGAAATTAAGAGACTTGACAA 5532
1016 sLeuThrThrGluThrGlyLysAsnAspTyrMetAlaLeuLeuProPhe 1033
|||||
5533 ATTAGCTTAGAGACCGGTGGAAGAAAGATGGGTGACCTCTCTCCCTTAG 5582
1033 aLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1049
|||||
5583 CCGTCTTAGAGGCGCAGAAATACCCCTGGCGGTTTGGTTAACTCTTAT 5632
1050 GluLeuLeuTyrGlyLysProProLeuAlaGluIleAlaPheAlaHis 1066
|||||
5633 GAAATTTCTATGAGAGCACCCCCATACTTGAATGCTGAGAAACTTT 5682

```



```

1066 sSerAlaAspValIleuLeuSerGlnProIleuPheSerArgLeuLysAlaL 1083
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5683 GGGTCCCGATGATAGATTCTCCCTGCTTATTACTCACTTAAGGCTT 5732
1083 euGIuTrpValArgIleuAlaIleuLysGlnLeuArgGluAlaIleuLys 1099
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5733 TAGAAATGTAGAGACCCAAATCTGGACCATCAAGAGGAGTATAG 5782
1100 G1yG1yAspLeuGlnValProHisArgPheGlnValG1yAspSerValTy 1116
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5783 CCGGTACCGTACACATCCCTCAACCCGCTTCAGTGGGGATCAAGTGC 5832
1116 ValArgArgHisArgAlaG1yAsnLeuGluTrpArgTrpLysG1yProT 1133
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5833 TGTCAGACGCCATCGACCCAGACGCTTGAGCTCGTGGAAGGCGCAT 5882
1133 yTrpLeuValIleuLeuThrThrProThrAlaValLysValGluG1yLeuPro 1149
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5883 ACCGTGGTGTGCTGACTACCCCGACGCGGTAAAGTCGATGGTATTGCT 5932
1150 ThrTrpIleuHisAlaSerHisValLysProAlaProPro.....ProAs 1164
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5933 GCGGTGGTCCATGCTCTCACTCAACCAACGACACCTTCGAGACACAGA 5982
1164 pSerG1yTrpArgAlaGluLysThrGluAsnProLeuLysLeuArgLeuH 1181
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5983 TGAGTCTCGGAGCTGGAAAGACTGATCATCTTAACTCGATATTC 6032
1181 IsArg 1182
6033 GGGCG 6037

seq_name: /cgn2.6/ptodata/2/ina/5A.COMB.seq:us-08-258-420-13

seq documentation block:
; Sequence 13. Application US/08258420
; Patent No. 5710037
; GENERAL INFORMATION:
; APPLICANT: Nienhuis, Arthur W.
; TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Vecto
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gillfillian, Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 Inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,420
; FILING DATE: 10-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 271010-208
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8202 bases
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: viral genome
US-08-258-420-13

alignment_scores:
  Quality: 3993.50      Length: 1210
  Ratio: 3.982          Gaps: 11
  Percent Similarity: 82.893      Percent Identity: 62.727

alignment block:
US-09-171-553b-5 x US-08-258-420-13/rev ...

Align seg 1/1 to reverse of: US-08-258-420-13 from: 1 to: 8202

1 G1yArgArgG1ySerAspProLeuProGluProArgValThrLeuLysVal 17
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5529 GGAAGTCAGAGGTCAAGAGCCGCCCTCGAACCAGATTAACCTCAAGT 5480
17 1GluG1yGlnProValGluPheLeuValAspThrG1yAlaGluHisSerV 34
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5479 CGGGGGCAACCCGTCACCTTCCTGTAGATCTGGGGCCCAACTCCG 5430
34 1LeuLeuGlnProLeuG1yLysLeuLysG1yLysSerTrpValMet 50
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5429 TGCTGACCCCAAAATCTCGAACCCTTAAGATTAAGTCTCGTGGGCCAA 5380
51 G1yAlaThrG1yGlnArgIleuTrpTrpThrThrArgArgThrValAs 67
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5379 GGGGCTACTGAGAGCAAGCGGTATCGCTGACACGATCGCAAGTACA 5330
67 pLeuG1yValG1yArgValThrHisSerPheLeuValIleProGluCysP 84
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5329 TCATGCTACCGGTAAAGTCAACCCCTTTCCTCCATGTACACAGACTGTC 5280
84 roValProLeuLeuG1yArgAspLeuLeuThrLysMetG1yAlaGluLe 100
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5279 CCAATCTCTGTTAGAGAGATTTCTGCTAACTAAATAAGCCCAATC 5230
101 SerPheGlu...GlnG1yArgProGluValSerValAsnLysProIle 116
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5229 CACTTTGAGGATCAGACGCTCAGTTATGAGCAATGGGCGACCCCT 5180
116 eThrValIleuThrLeuGlnLeuAspGluTrpArgLeuTyrSerProG 133
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5179 GCAAGTGTGACCTTAATATAGAGATGATTCGGCTACATGAGAACCT 5130
133 1nValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5129 CAATAAGAGCCAGATGTTCTTAGGGTCCACATAGGCTGTGATTTCT 5080
149 GlnAlaTrpAlaGluTrpAlaGlyMetG1yLeuAlaLysLysIleuProP 165
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5079 CAGGCTGGGGCGGAACCGGGGATGGAGCTGGCACTTGGCAAGCTCC 5030
165 oGluValIleGlnLeuLysAlaSerAlaThrProValSerValArgGln 182
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5029 TGTGATCATACCTTGAAGCAACCTCTACCCCGTGTGATTAACAAT 4980
182 yTrpLeuSerArgGluAlaArgGluG1yIleTrpProHisValGlnArg 198
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4979 ACCCATGTCAACAAGAGCCAGATGGGATCAAGCCCAACAACACAGA 4930
199 LeuIleGlnGlnG1yLeuValProValGlnSerProTrpAsnTrp 215
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4929 CTGTTGACCAAGGCTGACTGATCCCTCGAGTCCCTCGAAGACGCC 4880
215 oLeuLeuProValArgLysProG1yThrAsnAspTyrArgProValGln 232
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4879 CCTGCTACCGTTAAGAAACAGGAGCTAATGATTATAGCCGTGTCAG 4830

```

```

232 splenArgLysValAsnLysArgValGlnAspIleHisProThrValPro 248
4829 ATCTGAGAGAGTCAACAAGCGGTGAGAGATCCACCCACCGTGGCC 4780
249 AsnProThrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrTh 265
4779 AACCTTACACCTCTTGAGCGGGCTCCACCGTCCACAGTGGTACAC 4730
265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThr 282
4729 TGTGCTGATTAAAGAGATGCTTTTCTGCTGAGACTCCACCCACCA 4680
282 eGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
4679 GTACACCTCTCTCCCTTGAGTGAAGAGATCCAGAGATGGAAATCCA 4630
299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
4629 GGACATTTGACCTGGACCGAGCTCCACAGGGTTCAAAAACAGTCCAC 4580
315 rIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGln 332
4579 CCTGTTGATGAGGAGCTGCACAGAGACTGACAGACTCCGATCCAGC 4530
332 iSPROGlnValThrLeuLeuGlnTyrValAspAspLeuLeuValGly 348
4529 ACCAGACTTGATCCTGCTACAGTACGTGATGACTTACTGCTGGCCGC 4480
349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuGlu 365
4479 ACTTGTAGCTAGCTAGCTCCACAAGGTATCGGGCTGTGTTACAAACCT 4430
365 uSerAspLeuGlyTyrArgAlaSerAlaLysAlaGlnIleCysArg 382
4429 AGGGAACCTCGGTATCGGGCTCGGCCAGAAAGCCCAATTGGCCAGA 4380
382 rGlnValAlaThrTyrLeuGlyTyrSerLeuArgGlyGlnArgTrpLeu 398
4379 AACGGGTCAAGTATCTGGGGTATCTTAAAGAGGGTCAAGATGGCTG 4330
399 ThrGlnAlaArgLysLysThrValAlaGlnIleProAlaProThrAla 415
4329 ACTGAGGCCGAGAAAGAGACTGTGATGGCGGAGACTTCCGAAAGCCCC 4280
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrp 432
4279 TCCACAACTAAGAGAGTCTCTAGGAGCGAGGCTCTGCGCTCGGA 4230
432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
4229 TCCCTGGGTTGCAAAATGGCAGCCCTGTACCTCTCACCAAAAGC 4180
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAla 465
4179 GGGACTCTGTTTAATTTGGGGCCAGACACCAAAAGCCCTATCAAGAAAT 4130
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThr 482
4129 CAAGCAAGCTCTTCAACTGCCAGCCCTGGGTTGCCAGATTGGACTA 4080
482 ySPROpHeThrLeuTyrValAspGluArgGlyValAlaArgGlyVal 498
4079 AGCCCTTTAACTCTTTGTCGACGAGAAAGCAGGGCTACGCCAAAGCGCTC 4030
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLys 515
4029 CTAAACGCAAAAGCTGGACCTTGGCGTGGCGCGGTGGCTTACCTGTAA 3980
515 sLysLeuAspProValAlaSerGlyTyrProValCysLeuLysAlaIle 532
3979 AAAGCTAGACCCAGTGGCAGCTGGCGCCCTGCTACGATGGTGG 3930
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3929 CAGCATTTGCACTTCTGACAAAGATCTGTGCAAGCTCACTATGGACAG 3880
549 AsnIleThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnPr 565
3879 CCGTGTGTCATCTGGGCCCCCATGCGCTGAGAGCACTAGTTAAGCAACC 3830
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeu 582
3829 CCGTATGCTGGCTGCTCCAAATGCGGAGTACCCATTACCAAGCCCTGC 3780
582 eLys...ThrGluArgValThrPheAlaProProAlaLeuAsnPro 597
3779 TCTGAGACAGGACCGGGTCCAGTCTGGGCTAGTGGCCCTTAATTCAC 3730
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysH 614
3729 GCTACGCTGCTCCT...CTGCTGAGAGAGGGCTCCAAATGACTGCT 3683
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
3682 TGACATCTTGCTGCTGAGCCAGCACTAGATCAGATCTTAGGAGCCAC 3633
631 rLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrVal 647
3632 CCTCCCAAGACGCCAGCACACCTGTGTACAGATGGAGGAGCTTCCTG 3583
648 ValGluGlyLysArgMetAlaGlyAlaValAlaValAspGlyThrArgTh 664
3582 CAAGAAGGGCAGCGTAAGCGCGGAGCAGCGGTGACCACTGAGACTGAGGT 3533
664 rIleThrAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 681
3532 AATCTGGGCGAGGATGCTCCAGCGGAGCATCGGCCCAAGAGCTGAC 3483
681 eMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
3482 TCATAGCGCTCACCAAGCCCTAAAGATGGCAGAAAGTAAAGCTTAAT 3433
698 rIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyVal 714
3432 GTTATACATGATAGCGCTTACGCTTTTCCACCGCCCTATTCATGAGCA 3383
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluLys 731
3382 AATATACAGAAAGCGCGGGTGTCTACATTCAGAAAGAAAGATCAAGA 3333
731 sLysGluGluIleLeuSerLeuLeuGlnAlaLeuHisLeuProLysArg 747
3332 ACAAGAGAGATCTTACCTCTTAAAGGCTCTCTTGTGCCCAAAAGA 3283
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIle 764
3282 CTTAGCATATTCATTCGCCGCGGACATCAAAAGAAAGACGCGAGAGGC 3233
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGln... 779
3232 CAGGGGCAACCGGATGGCCGAGCAAGCGCCGAGAGTGGCCACTAGAG 3183
780 .....AlaValAsnLeuLeuProIleIleGluThrProLysAla 792
3182 AAATCCAGAGACTTCCACACTTGTGATAGAAATCTCAACCCCTATACC 3133
793 ProGluProArgArgGlnTyrThrLeuGluAspTrpGlnIleLysLys 809
3132 CATGAA...CACTTTCATACAGTACAGACAAAGATTTGACCA 3086
809 sIleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSer... 824
3085 ACTA.....GGAGCCACTTATGACAGCGCA 3060
825 .....TyrGlyLysGluIleLeuProHisLysGluGly 835

```

```

1132 ProTyrLeuValLeuLeuThrThrProThrAlaValValGluGlyIle 1148
1148 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2109 CCCTATACCGTCTGGTGAACACCCCCACCGCTCTCAAGTAGACGGCAT 2060
1148 eProThrTriPLeuHisAlaSerHisValValProAlaProProProAspS 1165
2059 ||::||| ||||| ||||| ||||| ||||| ||||| |||||
CGCTGGGTGATGCCACGGCTGCACGTAAGAGGGGGGACCAACCCCTCGG 2010
1165 eArgLy.....TTPaArgIaGluLyThrGluAsnPro 1175
2009 CCGGAACACACATCAGCAGCAGCATGGAGGTCCAGCGTTCTCAAAACCC 1966
1176 LeuLysLeuArgLeuHisArgLeuValPro 1185
1959 TTAAGATTAAGATTATTAACCCGTGGGGCCCCC 1930
seq_name: /cgm2_6/plotdate/2/ina/5A_COMB.seq:us-08-110-300A-8

```

```

seq.documentation_block:
? Sequence 8, Application US/08110300A
? Patent No. 5643756
? GENERAL INFORMATION:
? APPLICANT: Pinter, Abraham
? APPLICANT: Kayman, Samuel
? TITLE OF INVENTION: FUSION GLYCOPROTEINS
? NUMBER OF SEQUENCES: 25
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Davis Hoxie Faithfull and Hapgood
? STREET: 45 Rockefeller Pl.
? CITY: New York
? STATE: New York
? COUNTRY: USA
? ZIP: 10111
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/110.300A
? FILING DATE: 20-AUG-1993
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Jacobs, Seth H.
? REGISTRATION NUMBER: 32,140
? REFERENCE/DOCKET NUMBER: 11698A50
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-757-2200
? TELEFAX: 212-586-1461
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8323 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
?
? US-08-110-300A-8
?
alignment_scores:
? Quality: 3980.00 Length: 1206
? Ratio: 3.992 Gaps: 12
Percent Similarity: 82.670 Percent Identity: 62.687
alignment_block:
? US-09-171-553B-5 x US-08-110-300A-8 ..
Align seg 1/1 to: US-08-110-300A-8 from: 1 to: 8323
1 G1VARGRGGLYSERISAPROLEUPROG1UAROARGVALTHREULYSVA 17

```

2236 GGAGCTGAGGTGAGAGCCCGCCCTGAAACCCAGATACCTCAGAGT 2285
17 1GluGly6InProVal1GluPheLeuValAspThrGlyAlaIleHisSerV 34
2286 CGGGGGGCAACCCGTGACCTTCTAGTGAATGAGGGGCCCAACCTCG 2335
34 alleuLeuGInProLeuGlyLeuLysGluLysLysSerThrValMet 50
2236 TGTGACCAACCAATCTGGACCCCTAGTACAGACAGTCTCCGGGTCCA 2385
51 GlyAlaThrGlyAlaArgIleTyrProThrThrArgArgThrValAs 67
2386 GGGGTACTAGTGAAGGCGGTATCCCTGGACACGAGATCGCGCAGTCA 2435
67 PleuGlyVal1GlyArgVal1ThrHisSerPheLeuVal1LeuProGluCysP 84
2436 CTAAGCCACCGGTAAAGTCAACCATCTTCTCCATGATGACAGATTGCC 2485
84 rovalProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
2486 CTAATCCTCTGCTAGAGAGATTTGCTGACTAACTAAAGCCCAATTT 2535
101 SerPheGlu...GlnGlyArgProGluValSerValAsnLysProIle 116
2536 CACTTGGAGGATCAGAGACTCAGGTGTGGGCCCAATGGACAGCCCT 2585
116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
2586 GCAAGTGCTGACCTTAACATGAGAGATGAGTATGGCTTACATACAGCT 2635
133 InValLysProAspGlnAspIleGlnSer...ThrPleuGluGlnPhePro 148
2636 CAAAGGGCAGATGTGCTCTAGGGTCCACATGGCTCTCTGATTTTCCC 2685
149 GlnAlaThrPalaGluThrAlaGlyMetGlyLeuAlaLysGlnValProP 165
2686 CAGGCTGGGCGAGAAACGGGGGCATGGGGCTGGCCGTGCCCAAGCTCC 2735
165 oGlnVal1IleGlnLeuLysAlaSerAlaThrProValSerValArgIleT 182
2736 TCTGATCAATACCTCTGAAGGCAACCTCTACCCCGTGTCCATTAACAAT 2785
182 yProLeuSerArgGlnAlaArgGluGlyIleTyrProHisValGlnArg 198
2786 ACCCATGTGCACAAAGAACCCAGACTGGGGATCAAGCCCAATACAGAGA 2835
199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProThrPasnThrP 215
2836 CTGCTGAGTACAGGGAATCTGTACCTGCCACGCCCTCCGGAACAGGCC 2885
215 oLeuLeuProValArgLysProGluTyrHisAspTyrArgProValGlnA 232
2886 CCTGTACCCCTTTAGAAACCGGGGACTAAATGATTAAGGCTGTCCAGG 2935
233 sPLeuArgGluValAsnLysArgValGlnAspIleHisProThrValPro 248
2936 ATCTGAGAGAGTCAACAAGCGGTGGAAGACATCCACCCACCGTGC 2985
249 AsnProTyrAsnIleLeuSerAlaLeuProProGluArgAsnTyrThr 265
2986 AACCTTACCAACCTCTGAGGGGCTCCACCGTCCACACGAGGTGTAAC 3035
265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
3036 TGTGCTTGACTTAAGAATGCTTTTCTGCTGAGACATCCACCCACCA 3085
282 eArgInProLeuPheAlaPheGluTyrArgAspProGluTyrGlyArgThr 298
3086 GTCAAGTCTCTTTCGCTTTGAGTGAAGATCAGAGATGGGAATCTCA 3135
299 GlyLeuLeuThrThrArgLeuProGlnGlyPheLysAsnSerProH 315
3136 GGAACATTAACCTGACACAGACTCCCGCAGGGCTTTCAAAAAAGTCCAC 3185
315 rIlePheAspGluAlaLeuHisArgAspLeuLysAsnPheArgIleGln 332
3186 CCGTTTGTATGAAGCCCTGCACAGGAGCTGCAGACTTCCGGATCCAGC 3235
332 1sProGlnVal1ThrLeuLeuGlnTyrValAspAspLeuLeuAlaGly 348
3236 ACCCAACCTGTATCTGTCCAGTATGTAGATGACTTACTCTGCGCCGC 3285
349 AlaThrLysGlnAspCysLeuGlnIleThrLysAlaLeuLeuLeuGluLe 365
3286 ACTTGTGACTGTACTGTCAACAAGTACGGGGGCGCTTTCACAAACCT 3335
365 uSerAspLeuGlyTyrArgAlaSerAlaLysAlaGlnIleCysArg 382
3336 AGGGGACCTCGGATATCGGGCGCTGGCCCAAGAAAGCCCAATTTGCCAG 3385
382 rGluVal1ThrTyrLeuGlnTyrSerLeuArgGlyGlnArgTyrPleu 398
3386 AACAGGTCAAGTATCTGGGTATCTTCTAAAGGGGTGACAGATGGCTG 3435
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
3436 ACTGAGGCCAAGAAAGACACTGTATGGGGAGGAGCTTACCGAAGACCC 3485
415 aLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTyrI 432
3486 TCGAACATTAAGGGGAGTTCCTAGGGAGGCGAGGCTTCTGCGCTCTGGA 3535
432 1eProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
3536 TCCCTGGGTTTGCAGAAATGACACCCCTTTGACCTTCCACCAAAAGC 3585
449 LysGlyGlyPheSerThrPalaProGlnHisGlnLysAlaPheAspAlaI 465
3586 GGGACTGTGTTGAGTGGGGGCCAGACCAACAAAGCCATCAACAGAGAT 3635
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrI 482
3636 CAAGAGGCTCTCTTAACGTGCCCTGGGATTTGCCAGACTTACATA 3685
482 ySPProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
3686 AGCCCTTGAACCTTTTGTGAGAGAGAGGAGGTACGCCAAAGGTGTC 3735
499 LeuThrGlnThrLeuGlyProTyrArgArgProValAlaTyrLeuSerY 515
3736 CTAACGCAAAACTGGGGCGCTTGGCGTGGCGGCTGTACTCTGTCGAA 3785
515 sLysLeuAspProValAlaSerGlyTyrProValCysLeuLysAlaIleA 532
3786 AAACGTAGACCCAGTGTGACAGGCGGTGAGGCGCCCTGCACAGGAGTGA 3835
532 1AlaIleValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlnI 548
3836 CAGGCATATCGCGTTTGCAGAAAGCGTGGCAAGCTCCACATGGGACAG 3885
549 AsnIleThrVal1LeaIleAlaProHisAlaLeuGlnAsnIleValArgGluP 565
3886 CCACTAGTACTTGTGGGCCCCCAATGCAGTAAAGGACACTGTAAAGCAAC 3935
565 oProAspArgTyrMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
3936 CCGTGAATCGGTGCTCTCAACCCCGAATGACCACTACAGGCTGTGC 3985
582 eLeuLeu...ThrGluArgVal1ThrPheAlaProAlaAlaLeuAsnPro 597
3986 TTCTGAGACAGGACCAAGTCCAGTGTGGACCAATGTGGCCCTTAACCCA 4035
598 AlaThrLeuLeuProGluGluThrAspGluProVal1ThrHisAspCysH 614
4036 GCTACGCTGTGCTCC...CTACCTGAGGGGGGCTGCAACATGATGCTGCT 4082


```
seq documentation block:
; Sequence 8, Application US/00866642
; Patent No. 5952474
GENERAL INFORMATION:
APPLICANT: Pinter, Abraham
APPLICANT: Kayman, Samuel
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 45 Rockefeller Plaza, Suite 2800
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/886,642
FILING DATE: 01-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/110,300
FILING DATE: 20-AUG-1993
APPLICATION NUMBER: 07/938,100
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hone, William J.
REGISTRATION NUMBER: 26,739
REFERENCE/DOCKET NUMBER: 07763/010002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212/765-5070
TELEFAX: 212/258-7291
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 8323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-886-642-8

alignment_scores:
Quality: 3980.00      length: 1206
Ratio: 3.992          Gaps: 12
Percent Similarity: 82.670 Percent Identity: 62.687

alignment_block:
US-09-171-553B-5 x US-08-886-642-8 ..
Align seg 1/1 to: US-08-886-642-8 from: 1 to: 8323

1 GlyAATGATGGlySeraSPProLeuPProGLuPtoArGVAlThrLeuLysVa 17
||| :::::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
2236 GGAGTCACAGGCCTGAGGAGCCGCCCTGAACCCAGCATACCCTCACAGT 2285
17 LGIuLYgInPProVALGIuPhLeuValAsPTnrgLYALeGluHisserV 34
| ||||||||||| ||||||||||| ||||||||||| |||||||||||
2286 CGGGGGGCAACCGGTACTCTCTTAAGTAGTACTGGGGGCCAACACATCCG 2335
34 allLeuDLeuInPProLeuGLyLysLeuLysGLuLysLysSerTPvaIMet 50
|||| | ||| ||| ::::::|||:::||:::||:::||:::||:::||:::||
2336 TGCTGACCCAAATCCTGGACCCCTTAAGTCAGACAGTCTGCTGGGTCAA 2385
51 GIYAlaTrnGlnArgLntYrProTrPrThrThrArgArqThrValas 67
||||| ||||||| :::::||| ||||||| ||| |||
2386 GGGGCTACTGGAGGAGGACGGTATCCTGTGACACAGGATCGCGAGTGCA 2435
67 PleugIyValGIyArGVAlThrHIsseRPhleuValIlleProGLucysP 84
```

[illegible]

```
3336 AGGGACCTCGGATATCGGCGCTCGGCCAAGAAAGCCAAATTTCACAGA 3385
3382 rggluValThrTyrleuGlyTyrSerleuArgGlyGlyGlnArgTrpLeu 398
3386 AACAGGTCAAGTATCTGGGATATCTCTTAAAGAGGGTCAAGATGGGTG 3435
3399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
3436 ACTGAGGCCAGAAAGAGACTGTGATGGGAGCGCTTACTCCGAGACCCC 3485
415 alyslValAlaArgLuphLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
3486 TCGAACACTAAGGAGGTTCTAGGAGCGAGGCTTGTGCGCTCTGGA 3535
432 lProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
3536 TCCCTGGGTTTCAGAAATGCGAGCCCTTGTACCTCTGCACCAAAAGC 3585
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAla 465
3586 GGGACTCTGTGTGAGTGGGCGCCAGACAGCAAAAGGCTTACCAAGAGT 3635
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValTrp 482
3636 CAAGAGGCTCTTAACTGCCCCGCTGCGGATTCAGACTTGACTA 3685
482 ySProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
3686 AGCCCTCGACACTTTTGTGACGAGAAAGAGGCGGTACGCCAAAGGTGC 3735
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSer 515
3736 CTAAACCAAAACTGGGCGCTGGCGCTGGCGGCTTACCTGTCTCAA 3785
515 slyslLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIle 532
3786 AAAGCTAGACCCAGTGGAGCTGGGCGCCCTTGTCTACGAGATGTAG 3835
532 lAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3836 CAGCCATCGCGCTTGTACCAAGAGCGTGGCAAGCTCACCATGGGACAG 3885
549 AsnIlePheValIleAlaProHisAlaLeuGluAsnIleValArgGln 565
3886 CCACTAGTCTCTGGCCCCCATGCAGTAGAGGCGCTACTTAGCAACACC 3935
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeu 582
3936 CCCTGATCGCTGGCTCCCAAGCCCGAATGACCCACTACAGCGCTCGC 3985
582 eLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
3986 TTCTGGACAGGAGCAGCGAGTCCAGTTCGACCAATAGTGGCCCTTAACCA 4035
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCys 614
4036 GCTAGCGTCTGCTCCT...CTACCTGAGAGGAGGCGCTGCACATGACTGCT 4082
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIle 631
4083 TGACATCTTGGCTGAAGCCAGCAAGCAATGACCATCTTACGAGCAGC 4132
631 rLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTrpVal 647
4133 CTCTCCAGAGCGTGAACACACTGTGTACACAGATGGAGAGCGACTTCTG 4182
648 ValGlyLysArgMetAlaGlyAlaValAlaValAspGlyThrArg 664
4183 CAAGAGGGGAGCGAGCGGAGGAGCAAGTACCAACCGAGACGAGGT 4232
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGln 681
4233 AGCTTGGGCCAAAGCACTGCGAGCGGACATCGGCCCAAGAGCTGAGT 4282
681 eumAlaIleuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
4283 TGATAGCGCTACCCAGAGCCCTTAAATGACAGAGTAGAAGCTGAAAT 4332
698 IleTyrThrAspSerArgTrpAlaPheAlaThrAlaHisValHisGly 714
4333 GTTTACACCGATAGCCGTATGCTTTGGTCCACGTCCCATATTCACGAGA 4382
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGlnIleLys 731
4383 AATATATAGAAAGCGCGGTGTGTACATGAGAAAGAAAGAAATCAAAA 4432
731 snLysGluGluIleLeuSerLeuGluAlaLeuHisLeuProLysArg 747
4433 ATTAGCGAGATCTTGCCCTACTGAAGGCTCTCTCGGCCAAAGAGA 4482
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIle 764
4483 CTTAGCATATTCATTCGCCGGACATCAGAAAGGAAAGCCCGGAGGC 4532
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaGlnAlaVal 781
4533 AAGGGCAACAGGATGGCGCCAGCGGCCGAGAGTAGCCACTAGA. 4581
781 aLAsnLeuLeuProIleIleGluThrProLys..... 791
4582 .....CAACTCCAGAGACTTCACACTTCGTGATA 4611
792 .....AlaProGluProArg.....GlnTyrThrLeuGluAs 803
4612 GAAATTCAGCCCTTACTACTCATGACATTTTGCATATAGGAGCTGA 4661
803 pTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGly 820
4662 CATAAAGATGTGACTTAAGTAGGGCCACTTAGACATGCAAG...A 4708
820 hCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
4709 AGTGTGGGTATTCAGGGAAGCCGTGAATGGCTGATTCATTCACCTTT 4758
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrHisLe 853
4759 GAACATATGAGTTCTTCATCATATGACCCACCTCTGCTCTCAAAAAC 4808
853 uGlnGlnLeuValArgThrSer.....ProThrHisValLeuArgLeu 868
4809 AAAGCTCTTTCAGAAAGAACTACTGTCTTATTCATATGCGAGACCGG 4858
868 rGlyValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
4859 ATCGAAGCGTCAAGACATCACTGAGACTTGGCAAGCCGTGCACAGGTC 4908
885 AsnAlaAsnProSerArgTrpLeuProProGlyLysArgLeuArgLys 901
4909 AATGCGAGAGTCTGCGCTCAACAGAGGAGCTAGATTCGAGGCGACCG 4958
901 sProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLys 918
4959 ACCGGCACCACCTGGGAATGATTTCACTGAGGTAAGAAACCTGGCTGT 5008
918 yGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpVal 934
5009 ATGGTATTAATATCTTTAGTTTCATGACACTTCTCTGAGATGGGTA 5058
935 GluAlaTyrProThrLysGluThrSerThrValAlaLysLysIle 951
5059 GAAGCTTTCACCAAGAAAGAACTGCAAGTGTATACCAAGAGCT 5108
951 eLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySer 968
5109 ACTAGAAAGAAATCTTCCAGATTCGCAATGCGACAGGATTTGGGAACCG 5158
```


968 spasnqlyProAlaPheValAlaGlnValSerGlnIleuAlaLysIle 984
 |||||
 5159 ACATGGGCGCTGCTTGGTCTCCAGAGTAGTCAGACATCCCATTTA 5206
 985 LeuGlyIleAspTrpLysIleuHisCysAlaIaYarProGlnSerSerG 1001
 CTGGGGGTGATTTAGGAACACTACATGTGTGGCTTACAGACCCCAAGTTTCAG 5255
 1001 yGlnValGlnuArgMetAsnArgThrIleLysGluThrLeuThrLysLeu 1018
 |||||
 5259 TCAGGTGAAAGAAATGAAATGAGACATCAAGAGAGACTTAACTAAATATGA 5308
 1018 hrThrGlnTrpGlyIleAsnAspTrpMetAlaIleuLeuProPheValI 1034
 |||||
 5309 CGCTTGCACTGGCTCTAGGAGCTGGGTCTCTGCTTCCCTTCCCTACCTGT 5358
 1035 PheArgValAlaArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlu 1051
 |||||
 5359 TATCGAGCCCGCAACAGCGCGGGGCCCGCCATGTGTCTACCCCATATGAAT 5408
 1051 uLeuTyrGlyGlyProProProLeuAlaGlnIleAlaPheAlaHisSerA 1068
 |||||
 5409 CTATTATGGGCGACCGCCCGCCCTGTGAACCTTCCTGATCCGTACATGG 5458
 1068 IaAspValIleuLeuSerGlnProLeuPheSerArgLeuLysAlaIleuGlu 1084
 |||||
 5459 CAAGGTACTACTATACCCCTCTCTCCAAAGCCCATTTACAGGCACTCTAC 5508
 1085 TrpValAlaArgIlaArgAlaIaTrpLysGlnLeuArgIleAlaIaYrSerGly 1101
 |||||
 5509 CTGGTCCACACAGAACTGTGGAGACCGTGGGGGAGCGTTACCAA...GA 5555
 1101 yAspLeuGln.....ValProHisArgPheGlnValGlnLysSerV 1115
 |||||
 5556 AACACGTGACCGCGCGGTAGTGCTCTCACCCTTTCCGAGTGGGTATCACAG 5605
 1115 aIaYrValaIaArgIlaArgIlaArgIlaGlnLeuGlnIleThrArgTrpLysGly 1131
 |||||
 5606 TGTGGGTCCGACAGACCAACTAAATCTAGAACCCCGGTGAAAGGA 5655
 1132 ProTyrLeuValLeuLeuThrThrProThraAlaValLysValGlnGlyI 1148
 |||||
 5656 CCGTATACCGTCTTACTAGACACCCCAACCCCTCTCAAGGTGAGCGCAT 5705
 1148 eProThrTrpLleHisAlaSerHisValLysProAla..... 1160
 |||||
 5706 TGCACGCGTGAATCCACGCTGCCACGTAAAGGCTGCGGACACACAGATTGG 5755
 1161 ..ProProProAspSerGlyTrpArgAlaGlnIleThrGlnAsnProLeu 1176
 |||||
 5756 AACCCACCAATCGGAATCGACATGGCGGTGTCGAACGCTCTCAAAATCCCTTA 5805
 1177 LysLeuArgLeuHisAsnArg 1182
 |||||
 5806 AAGATTAGATTGACCGCG 5823

```
seq_name: /cgn2_6/ptodata/2/ina/PCrtus_COMB.seq: PCT-US93-08041-8
seq_documentation_block:
  : Sequence 8, Application PC/TUS9308041
  : GENERAL INFORMATION:
  : APPLICANT: The Public Health Research Institute of the City of
  : APPLICANT: New York, Inc.
  : TITLE OF INVENTION: FUSION GLYCOPROTEINS
  : NUMBER OF SEQUENCES: 16
  : CORRESPONDENCE ADDRESSES:
  : ADDRESSEE: David Hoxie Faithfull and Hapgood
  : STREET: 45 Rockefeller Pl.
  : CITY: New York
  : STATE: New York
  : COUNTRY: USA
  : ZIP: 10111
  : COMPUTER READABLE FORM:
  :
```

```

1 MEDIUM TYPE: FLOPPY DISK
2 COMPUTER: IBM PC COMPATIBLE
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: PatentIn Release #1.0, Version #1.25
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: PCT/US93/08041
7 FILING DATE:
8 CLASSIFICATION:
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 07/938,100
11 FILING DATE: 28-AUG-1992
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: US 08/____
14 FILING DATE: 20-AUG-1993
15 ATTORNEY/AGENT INFORMATION:
16 NAME: Jacobs, Seth H.
17 REGISTRATION NUMBER: 32,140
18 REFERENCE/DOCKET NUMBER: 11698A50
19 TELECOMMUNICATION INFORMATION:
20 TELEPHONE: 212-757-2200
21 TELEFAX: 212-586-1461
22 INFORMATION FOR SEQ ID NO: 8:
23 SEQUENCE CHARACTERISTICS:
24 LENGTH: 8323 base pairs
25 TYPE: nucleic acid
26 STRANDEDNESS: double
27 TOPOLOGY: linear
28 MOLECULE TYPE: DNA (genomic)
29 HYPOTHEICAL: NO
30 ANTI-SENSE: NO
31 PCT-US93-08041-8

```

Quality:	3980.00	Length:	1206
Ratio:	3.992	Gaps:	12
Percent Similarity:	82.670	Percent Identity:	62.687

```
alignment_block:
```

Align seg 1/1 to: PCT-US93-08041-8 from: 1 to: 8323

```

1  GLAFArGlsySertspProLeuPProGluPProAlqValThLeuLysVa 17
2236  GGGAGCTCAGGGGTCAAGAGCCGCCCTTGTGAACCAAGATTAACCTCAGAGT 2285
17  IGtugLyGlnProValGlpheLeuValAspThrGlyAlaGluHISser 34
2286  CGGGGGGGCAACCCGTCACCTCTCTAGTGGATACTGGGGCCCAACACTCG 2335
34  aLeuLeuGlnProLeuGlyLysLeuLysGluLysSerPraiMet 50
2336  TGTGTACCAACAATCTGTGACCCCTTAAGTACAACTGTGCTGGGTCCAA 2385
51  GLyAlaThrGlyGlnArgGlnTyrProThrThrArgArgThrValAs 67
2386  GGGGCTACAGGAGGGAGAGGGGTATGTGTGGACACAGCATGCCGAGTGC 2435
67  pleuGlyValGlyArqValThHisSerPheLeuValIleProGluLysP 84
2436  CCTAGCCACGGSTAAAGTACCCACTTCTTCTTCCTCATTCACCAATTGCC 2485
84  roValProLeuLeuGlyArqAspLeuLeuThrLysMetGlyAlaGluIle 100
2486  CCTATCCCTCTGTAGGAAGAAGTTTGCTGACTTAACATAAAGCCCAAAATT 2535
101  SerPheGlu..GlnGlyArqProGluValSerValAsnAsnLysProII 116
2536  CACTTTAGGGGATCAGAGCTCAGGTTCTGGGACCACATGGACACAGCCCT 2585
116  eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133

```

2586 GCAAGTCTGACCTAAACATAGAAGATGAGTATCGGCTACATGAGACCT 2635
 133 InVallyProAspGlnAspIleGlnSer...TriLeuGlnGlnPhePro 148
 2636 CAAAAGGGCCAGATGCTGCTAGGCTCCACATGGCTCTCTGATGATTTTCCC 2685
 149 GlnAlaTrpAlaGlnThrAlaGlyMetGlyLeuAlaLysGlnValaProPr 165
 2686 CAGGCGCTGGCCGAGAACCGGGGCTGGGCTGGCCGTGCGCCAGCTCC 2735
 165 oclInValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTr 182
 2736 TCTGATCATCTCTGAAGGACACCTACCCCGGTGCTCAATAAACAT 2785
 182 yTrProLeuSerArgGlnAlaArgGlyIleTrpProHisValGlnArg 198
 2786 ACCCATGTCCAGAGAGCCAGACTGGGGATCAAGCCCATACAGAGGA 2835
 199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPr 215
 2836 CTGCTGGATCAGGGAATCTGTGCTACCCGCAAGTCCCTGGAACAGCC 2885
 215 oLeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGln 232
 2886 CCGCTACCCGCTTAAGAAACCGGGAGCTAATGATATAGCCCTGTCAGG 2935
 232 sPLeuArgGlnValaLysLysArgValGlnAspIleHisProThrValPro 248
 2936 ATCTGAGAGAGTCAACAGCGGGTGAAGACATCCACCCACCTGCTGCC 2985
 249 AsnProTyrAsnLeuLeuSerAlaLeuProProGlnArgAsnTrpTyrTh 265
 2986 AACCTTACACCTCTTGAGCGGGCTCCACCCGTCCACACATGGTTACAC 3035
 265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
 3036 TGTGCTGACTTAAAGATGCTTTTCTGCTGAGACTCCACCCACACA 3085
 282 erGlnProLeuPheAlaPheGlnTrpArgAspProGlyThrGlyArgThr 298
 3086 GTGAGTCTCTTCCGCTTGAAGTGAAGATCCAGACATGGGAATCTCA 3135
 299 GLyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
 3136 GGACAAATTAACCTGGACCACTCCCGCAGGTTTCAAAAACAGTCCAC 3185
 315 rIlePheAspGlnAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
 3186 CCTGTTGATGAAGCCTCGACAGGACCTCGCAGACTTCGCGATCCAGC 3235
 332 tSProGlnValThrLeuLeuGlnTrpValaLysAspLeuLeuAlaGly 348
 3236 ACCCAGACCTGATCTGCTCCAGATATAGATCTACTGCTGCCGCC 3285
 349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuGlu 365
 3286 ACTTGTGAGCTTGACTGTCAACAGATAGCCGGGCTGTTACAACACCT 3335
 365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
 3336 AGGGGACCTGGATATCGGGCTCGGCCAGAGAAAGCCCAATTTGCCAGA 3385
 382 rGlnValaThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
 3386 AACAGGTCAAGTATCTGGGGATCTTCTTAAAGAGGGGTCAAGATGGCTG 3435
 399 ThrGlnAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
 3436 ACTGAGGCCAGAAAGAGACTGTGATGGGCACGCTACTCCAGAGACCCC 3485
 415 aLysGlnValaArgGlnPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
 3486 TCGACAACTAAGGAGTTCCTAGGGACGGCAGGCTTGTGCGCTCTGGA 3535

4432 lProGlyPheAlaThrLeuAlaAlaLarProLeuTyrProLeuThrLysGlu 448
 3536 TCCCTGGGCTTTGAGAAATGGCACCCCTTTTATCCCTCTCCACAAACG 3585
 449 LysGlyGlyPheSerTrpAlaProGlnHisGlnLysAlaPheAspAlaI 465
 3586 GGGACTGTGTTGATGGGGGCCAGACCCAGACCAAGAAAGGCTCAACAGAT 3635
 465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValaThr 482
 3636 CAGAGAGGCTCTCTTAAGTCCCGCTGCGCTGGATTCAGACTGTGACTA 3685
 482 ySProPheThrLeuTyrValaLysAspLysArgGlyValaAlaArgGlyVal 498
 3686 AGCCCTTGCAACTTTTGTGACGAGAGCGAGGCTACGCCCAAGAGGTGTC 3735
 499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrIleuSerLy 515
 3736 CTACCGCAAAACCTGGGCTTTGGCGTGGCGCGGTGCGCTACGTCCAA 3785
 515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIle 532
 3786 AAGCTAAGACCAAGTGGAGCTGGGTGGCGCCCTTGCTACGATGTGTAG 3835
 532 lAlaValaAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
 3836 CAGCCATCCCGCTTGACGAAAGAGCGTGGCAACCTCACCCATGGGACAG 3885
 549 AsnIleThrValIleAlaProHisAlaLeuGlnAlaSerIleValArgGlnPr 565
 3886 CCACATGATCTTTGCGCCCGCCATGACAGTGAAGGACATGATTAACCAAC 3935
 565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
 3936 CCTGATCGCTGGCTCTCCACAGCCCGCAATGACCACTACCGGCTGTGC 3985
 582 eLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
 3986 TTTGGACAGGACCGAGTCCAGTCTGGACCAATAGTGGCCCTTAACCCA 4035
 598 AlaThrLeuLeuProGlnGlnThrAspGlnProValThrHisAspCysHis 614
 4036 GCTACGCTGTCTCCT...CTACCTGAGAGAGGCTGCAACATGACCTGCT 4082
 614 sGlnLeuLeuIleGlnGlnThrGlyAlaArgLysAspLeuThrAspIleP 631
 4083 TGACATCTTGCTGAGAGCCCGCAGGAACTAGACCAATCTTACGAGCAC 4132
 631 rOleuThrGlyGlnValaLeuThrTriPheThrAspLysSerSerTyrVal 647
 4133 CTCTCCAGACGCTGACACACCTGTGTACCAATGGAGACACCTTCGCG 4182
 648 ValGlnGlyLysArgMetAlaGlyAlaAlaValaValaLysPlyThrArgTh 664
 4183 CAAGAGGGGCGCGCCAGAGGCGGAGCAGCAGTAAACCCAGAGCAGAGGT 4232
 664 rIleTrpAlaSerSerLeuProGlnGlyThrSerAlaGlnLysAlaGlnu 681
 4233 AGTGTGGCCAAAGCACATGCGCAGCGGACATCGGCCCAAGAGGTGACT 4282
 681 eMetAlaLeuThrGlnAlaLeuArgLeuAlaGlnGlyLysSerIleAsn 697
 4283 TGAATCGCTCAACCCAGAGCTTAAATAAGCAGAGAGTGAAGAGTCAAT 4332
 698 lIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
 4333 GTTTCACCGAATAGCCGTTATGCTTTGCGCCTCCATCCATTCACGGAGA 4382
 714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGlnIleLysA 731
 4383 AATATATAGAAAGCGCGGGTGTGCTACATCAAGAGGAAGAAATCAAAA 4432

```

731 snlysgluGluIleuSerIleuGluAlaIleuHisIleuProIysArg 747
|||||
4433 ATAGAGCAGAGATCTTGCCCTACAGAGCTCTCTCTGCGCAAGAA 4482
748 LeuAlaIleIleHisCysProGlyHisGluAlaIleAspIleIle 764
|||||
4483 CTAGAGTAATTCATTCGCGGACATCGAAGGAAACCGCGGAGAGC 4532
764 rArgGlyAsnGlnMetAlaAspArgValAlaIleGlnAlaIleVal 781
|||||
4533 AAGGGCAACAGAGATGGCCGACCAAGCGCCGAGAACTGACCTAGA. 4581
781 alaIleuLeuProIleIleGluThrProIys..... 791
|||||
792 .....AlaProGluProIleArg.....GluThrIleuGln 803
|||||
4612 GAAATTCAGCCCTATACATGAACTTCATATACGAGTCTG 4661
803 pThrGlnGluIleuLysIleAspGlnPheSerGluThrProGluGly 820
|||||
4662 CATAAAGATCTGACTAAACTAGGGCCACTTATGACGATGCAAG...A 4708
820 hrCysTyrThrSerTyrGlyLysGluIleuProHisLysGluGlyLeu 836
|||||
4709 AGTGTGGGTTCATCAGGAAAGCCGTAAATGCTGATCAATTCACCTTT 4758
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHis 853
|||||
4759 GACATATTAATTTCTTCATCAATTCACCCACTCAGTTTCTCAAAAAC 4808
853 uGlnGluLeuValArgThrSer.....ProTyrHisValIleuLeu 868
|||||
4809 AAAGCTCTCTCTGAAAGAACTAGCTCTTATTCATGCTGAAACGGG 4858
868 roGlyValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
|||||
4859 ATCGAAGCGTCMAAGACATCATGAGACTTGCACAGCTGTCGACAGGTC 4908
885 AsnAlaAsnProSerArgIleProProGlyLysArgLeuArgLysSer 901
|||||
4909 AATGCCAGCAAGTCTGCCGTCACAAAGAGACTAGAGTTGAGGACACG 4958
901 sProGlyAlaHisIleIleValAlaAspPheThrGluValLysProAlaLys 918
|||||
4959 ACCGGGACCCACTCGGAAATGATTCAGTGAAGAACTGCGCTGT 5008
918 yGlyAsnLysTyrIleuValPheValAspThrPheSerGlyThrVal 934
|||||
5009 ATGGGTATTAATTCATTTAGTTTCATAGACACTTCTCGAGGGTA 5058
935 GluAlaLysProThrLysGluThrSerThrValAlaIleLysLys 951
|||||
5059 GAAGCTTCCCAACCAAGAAAGAACTGCAAGTTGTAACCAAGAACT 5108
951 eleuGluGluIlePheProArgPheGlyIleProLysValIleGlySer 968
|||||
5109 ACTAGAAAGAAATCTCCAGATTCGCGATGCGACAGGTATTTGGGAACG 5158
968 spaSngIyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 984
|||||
5159 ACAATGGGCTCTCTGCTCTCCAGAGTACAGACGATGAGCGCATTTA 5208
985 LeuGlyIleAspTyrLysLeuHisCysAlaLysIleArgProGlnSer 1001
|||||
5209 CTGGGGGTGATTTGAAACTGATTCATGACGACCCAGAGTTGAGG 5258
1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeu 1018
|||||
5259 TCAGGTACAGAAATGAATAGACAAATCAAGAGACTTAACTAAATTTGA 5308
1018 hrThrGluThrGlyIleAsnAspTyrMetAlaIleuLeuProPheValLeu 1034
|||||
5309 CGCTTGCAACTGGCTCTAGGACACTGGTGTCTCTCTCCCTAGCCCTG 5358
1035 PheArgValArgAsnThrProGlyIlePheGlyLeuThrProTyrGlyLeu 1051
|||||
5359 TATCGAGCCCGCACACAGCGCGGCGCCCATGCTGATCCCATTTGAAT 5408
1051 uLeuTyrGlyLysProProIleuAlaGluIleAlaPheAlaHisSer 1068
|||||
5409 CTTATATGGGGCACCCCGCCCTTGTAACTTCCTGATCCTGACATGCG 5458
1068 IaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaIle 1084
|||||
5459 CAAAGTTACATCATTAACCTCTCTCCAGGCCATTTACAGGACTTAC 5508
1085 TrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGly 1101
|||||
5509 CTGGTCCAGCAGAAAGTCTGAGACCGTTGGCGGACCTTACCAA...GA 5555
1101 yAspLeuGln.....ValProHisArgPheGlnValGlyAspSer 1115
|||||
5556 ACAATGAGACCGCGCGGTAGTGCCTCACCTTCCGAGTGGAGACAG 5605
1115 aLTrpValArgArgHisArgAlaGlyAsnLeuGluThrArgTyrPlyGly 1131
|||||
5606 TGTGGTCCCGCAGACACCAAACTAAATCTAGAACCCCGCTGAAAGCA 5655
1132 ProTyrLeuValIleuLeuThrThrProThrAlaValLysValGly 1148
|||||
5656 CTTATACCGCTCTACGACTACCCCGCTCTCAAGTGGAGCGCAT 5705
1148 eProThrTrpIleHisAlaSerHisValLysProAla..... 1160
|||||
5706 TGCAGCTGATTCACGCTCCACGTAAGAGTGGCGACACAGAGATTG 5755
1161 ..ProProIleAspSerGlyTyrArgAlaGluLysThrGluAsnProLeu 1176
|||||
5756 AGCCACCATGGAGATGACATGCGCTGTTCAACGCTCAAAATGCCCTGA 5805
1177 LysLeuArgLeuHisArg 1182
|||||
5806 AAGATAAGATTGACCCG 5823
seq_name: /cgn2.6/ptodata/2/ina/5A_COMB.seq:us-08-110-300A-9
seq_documentation_block:
; Sequence 9, Application US/08110300A
;
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,300A
; FILING DATE: 20-AUG-1993
; CLASSIFICATION: 433
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H.
; REGISTRATION NUMBER: 32,140
; REFERENCE/DOCKET NUMBER: 11698A50

```



```

|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3860 CTACGCCAAACATGGGGCTTGGCGTGGCGCGGTGCTACCTGTCCAA 3811
515 slysluaspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3810 AAAGGTAGACCCAGTGGCAGCTGGGTGGCCCTTGCTTACGATGGTAG 3761
532 lalValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3760 CAGCCATGCCCTTCTTACCAAGAGCGCTGGCAAGCTCACCATGGAGAG 3711
549 AsnIleThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnPro 565
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3710 CCACATGATCATTTGGCCCCCATGGCATAGACAGCTAGTTAAGCAACC 3661
565 oProspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeu 582
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3660 CCGTATATGCTGGCTCTCCAAAGCGCCCAATGACCCATACAGAGCTCTGC 3611
582 euLeu...ThrGluArgValThrPheAlaProAlaAlaLeuAsnPro 597
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3610 TTCTGGACACGACGACGAGCTCAGTTCGGACCAATAGTGGCCCTTAAACCA 3561
598 AlathrLeuLeuProGluGluThrAspGluProValThrHisAspCysH 614
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3560 GCTACGCTGCTCCCT...CTACCTGAGAGAGGCGCTGCACATGACTCCCT 3514
614 sGlnLeuLeuIleGluGluThrGlyAlaArgLysAspLeuThrAspIle 631
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3513 TGACATCTTGGCTGAGAGCCAGCAGACCTAGACAGATCTTACGAGCAGC 3464
631 rGluLeuThrGluValLeuThrPheThrAspLysSerSerGlyVal 647
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3463 CTCCTCCAGACGCTGACACACCTGCTACACAGATGGAGCAGCTTCCG 3414
648 ValGluGlyLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgTh 664
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3413 CAAGAGGGGAGCGCAAGCGCGAGCAGTACACAGACCGAGACCGAGGT 3364
664 rIleThrAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 681
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3363 AGTCTGGGCCAAAGACACTGCCAGCGGAGCATGCGCCCAAGAGCTGAGT 3314
681 euMetAlaLeuThrGlnAlaLeuArgLysAlaGluGlyLysSerIleAsn 697
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3313 TGATAGCGCTCACCAAGCTTAAATAATGGCAGAGTAAGACCTGAT 3264
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3263 GTTACACCGATAGCCGTTATGCTTTTCCACCTGCCATATTCACGGAGA 3214
714 alleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3213 AATATAATAGAAAGCGCGGGTGTCTACATCAGAAAGAAAGAAATATCAAA 3164
731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3163 ATAAGAGAGAGATCTTGGCTCTACTGAAAGGCTCTTCTGCGCCAAAGA 3114
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIle 764
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3113 CTTACATTAATTCATTTGCCCGGAGCATCAGAGGAAAGCCGCGGAGGC 3064
764 rArgGlyLysGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaVal 781
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3063 AAGGGGCAACAGATGGCGCCAGCGGCGCGAGAAAGTAGCCACTAGA. 3015
781 alaAsnLeuLeuProIleIleGluThrProLys..... 791
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
3014 .....GAAAGCTCAGAGACTTCACACTTCTGATA 2985
792 .....AlaProGluProArgArg.....GlnTyrThrLeuGlnLys 803
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

```

2984 GAAATTCAGCCCTATATCATGACATTTTCAATATACGTGACTGA 2935
803 pThrGlnIleIleLysIleAspGlnPheSerGluThrProGluGly 820
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2934 CATAAAGATCTGACTAATAAGGGGCCACTTATAGCATGACAAAG...A 2888
820 hrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2887 AGTGTGGGTTTATCCAGAAAGCCTGAATGCTGTATCAATTCACCTTT 2838
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGluThrHisIle 853
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2837 GAACATATTAATTTCTTCAATTCATCCACCTCAGTTTCTCAAAAC 2788
853 uGlnIleLeuValArgThrSer.....ProTyrHisValLeuArgLeu 868
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2787 AAAGCTCTCTCAGAAAGAACTAGTCTCTTATTCATGCTTCAACCGG 2738
868 rGluValAlaAspSerValLysHisCysValProCysGlnLeuVal 884
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2737 ATCGACAGCTCAAAAGACATCATGAGCTTGGCCAAAGCTGCGACAGCTC 2688
885 AsnAlaAsnProSerArgIleProGlyLysArgLeuArgGlySerH 901
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2687 AATGCCAGCAAGTCTGCCCTCAAAAGAGAGACTAGAGTTCGAGGCGCCG 2638
901 sProGlyAlaHisThrGluValAspPheThrGluValLysProAlaLysT 918
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2637 ACCGGGACCCACTGGGAAATGATTTCACTAGTAAACCTGCGCTGT 2588
918 TyrGlnAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTyrVal 934
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2587 ATGGGATTAATATCTTTTATGTTTATGATGACACTTCTCTGGATGGTA 2538
935 GluAlaTyrProThrLysLysGluThrSerThrValAlaLysLysI 951
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2537 GAAGCTTCTCCCAACCAAGAAAGAACTGGCAAAAGTTGAACCAAGAACT 2488
951 GluGluGlnIlePheProArgPheGlyIleProLysValIleGlySerA 968
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2487 ACTAGAAAGAAATCTTCCCGAGATTCGGCATCGCACAGGATTTGGGAACG 2438
968 spAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIle 984
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2437 ACAAGGGGCTGCTCTCTCCATCCAAAGTAATGACAGACAGTACCGATTTA 2388
985 LeuGlyIleAspThrLysLeuHisCysAlaTyrArgProGlnSerSerG 1001
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2387 TTGGGGGTGATTTGGAACCTATGCTTGTACAGACCCAGAGTTTCAGC 2338
1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuT 1018
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2337 TCAGGTAAAGAAATGATATAGACATCAAGAGACCTTAACTAAATTTGA 2288
1018 hrThrGluThrGlyIleAsnAspThrMetAlaLeuLeuProPheValLeu 1034
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2287 CGCTTGAAGTGGCTTGAAGGATGGGTGCTCTGCTTCCCTTGCCTG 2238
1035 PheArgValAlaArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlu 1051
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2237 TATGAGCCGCCACACAGCGGGGCCCGCCAGCGTCTCAACCCCATATGAAT 2188
1051 uLeuTyrGlyGlyProProLeuAlaGluIleAlaPheAlaHisSerA 1068
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2187 CTATATATGGGACACCCCGCCCTTGAATCTCCCTGATCCSGATGAG 2138
1068 laAspValIleLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2137 CAAGAGTACTCATTAACCCCTCTCCAAAGCTCATATTACAGCAGCTTAC 2088
1085 TyrValArgLysAlaArgAlaTyrLysGlnLeuArgGluAlaTyrSerGly 1101
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
2087 CTGGTCCAGCAGCAAGTCTGGAAGCCGTTGGCGGACGCTTACCAA...GA 2041

```

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-9
seq_documentation_block:
Sequence 9, Application US/08886642
Patent No. 5952474
GENERAL INFORMATION:
APPLICANT: Pinter, Abraham
APPLICANT: Kayman, Samuel
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESS: Fish & Richardson P.C.
STREET: 45 Rockefeller Plaza, Suite 2800
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/886,642
FILING DATE: 01-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/110,300
FILING DATE: 20-AUG-1993
APPLICATION NUMBER: 07/938,100
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hone, William J.
REGISTRATION NUMBER: 26,739
REFERENCE/DOCKET NUMBER: 07763/010002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212/765-5070
TELEFAX: 212/258-2291
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 10367 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-886-642-9
alignment_scores:
```

Quality:	3960.00	Length:	1206
Ratio:	3.992	Gaps:	12
Percent Similarity:	82.670	Percent Identity:	62.687

```
alignment_block:
  no 00 177 553b
```

US-09-171-553B-5 x US-08-886-642-9/rev

Align seg 1/1 to reverse of: US-08-886-642-9 from: 1 to: 10367

1	GLYVALARGISERASERPROLEUPROGLUPROVALTHIRLEU	17
5360	GGAGGTCAGAGGTCAGGAGGCCCCCTGACACCCAGGATACCCCTAGAGT	5311
17	IGLUGLYNPROVALGLUPHELEUVALASPTHIRGLYLAQIUNHISERV	34
5310	CGGGGGGACACCCGTCACCTCTCTAGTGAGATACCTGGGCGCACACCTCG	5261
34	ALILEUENGINPROLEUGLYLSEULYSGLULYSSETTRVALMET	50
5260	TGCTGACCCCAAAATCCTGGACCCCTTAAGTAGACATCTGCTGGGTCCAA	5211
51	GLYALATHIRGLIHLAARGLINTPROTPRTHIRATAGAGTHVALAS	67
5210	GGGGCTACTGGAGGAGGAGGGATAGCTGCTGGACACGATCCGCCAGTCA	5161
67	PLEUGLYVALGLYATVALTHIRHISERPHLEUVALILEPROGLUCYSP	84
5160	CTTAGCACCGGTAAGGTCACCCATCTCTCTCCATGATACAGATATGCC	5111
84	ROYALPROLEULEUGLYATGASBPDEUETHIRLYSMETGLYALAGNILE	100
5110	CCATATCTCTGTGTAGGAAAGATTTGCTGACATAAACTAAAGCCCAATTT	5061
101	SETPHGLU... GLINGLYATPROGLIUALISERVALASNASLUPROII	116
5060	CACCTTGGAGGATAGGAGCTAGGTTGGTGGACCAATGGCAGACCCCT	5011
116	ETHVALLEUTHIRLEUGINLEUASPAPOIUNYATRGLEUTHIRSERPROG	133
5010	GCAAGTCGTGACCCCTTAACATAGAAAGATGATATCGGCTACATGAGACT	4961
133	INVALLYSPROASPGINASPILLEGINSER... TRPLEUGLINPHEPRO	148
4960	CAAAAGGCGCAGATGTCCCTCTAAGGTCACATGGCTCTCTGATTTTCC	4911
149	GLIALATHRALAGIUTHIRAGIYMETGILEUALIYGLINVALPROPR	165
4910	CAGGCTGGGACAAACCGGGGCGATAGGGGCTGGCCGTTCGCCAAGCTCC	4861
165	OGIVALIILEGINLEULYSALASERVALATHRPROVALSERVALARGINT	182
4860	TCGTGATCATACCTCTAAGGCAACCTTACCCCGCTGCCATTAACAAAT	4811
182	YRPROLEUSERAGIUALIARGGLUGLYLIEETRPROHISVALGINARG	198
4810	ACCCCATGTCACAAAGAACGCAACATGGGATCAACCCCATCATACAGAGA	4761
199	LEUILLGLINGLYLIELEUVALPROVALGINSETPROTPRASPETHR	215
4760	CTCTGTGATGAGGAATTTGTGTACCTGCCAGTCCCTCGGAACACCC	4711
215	OLEULEUPROVALARGIYSPROGLYTHIRASAPTYATGPROVALGINA	232
4710	CTGTGATCCCGTTAAACAAACCGGGAGCATATGATTATATAGGCTGTCCAGG	4661
232	SPLUENARGIUALASNLVSATGVALGINASPILEHISPROTHYALPRO	248
4660	ATCTGAGAGAGTCAACAAAGCGGTGGGAAGACATCCACCCACCGTCCGC	4611
249	ASNPOTGYASMEULEUSERIALLEUPROPOLIUATGASINTPTTYRTH	265
4610	AACCTTACAACTCTTGAGCGGGGCCACCGCTCCACCAATGGTACAC	4561

```

265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
|||||
4560 TGGCTTGAAGTAAAGATGCTTTTCTGCTGAGACTCCACCCACCA 4511
282 eGlnProLeuPheAlaPheGluThrParAspProGlyThrGlyArgThr 298
|||||
4510 GTGAGTCTCTCTTCCGCTTGTGAGTGGAGATCCAGAGATGGGAATCTCA 4461
299 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
|||||
4460 GGACATATTAACTGGACAGACTCCCGAGGCTTCAAAAACAGTCCAC 4411
315 rLePheAspGluAlaLeuHisArgAspLeuAlaSnPheArgLeuGln 332
|||||
4410 CCTGTTGATGAAGCCCTGCACAGGACCTCCAGACTTCGAGATCCAGC 4361
332 AspProGlnValThrLeuGlnGlyThrValAspAspLeuLeuAlaGly 348
|||||
4360 ACCAGACTGATTTCTGCTCCAGTATGTAGATGACTTACTGCTGCCGCC 4311
349 AlaThrLysGlnAspCysLeuGlyThrLysAlaLeuLeuLeuGluLe 365
|||||
4310 ACTTGTGACTTGAAGTCAACAGATGACGGCGCTGTTCACAAACCT 4261
365 userAspLeuGlyThrArgAlaSerAlaLysLysAlaGlnIleCysArg 382
|||||
4260 AAGGGACTCGGATATCGGCGCTCGCCCAAGAACCCCAATTTCCAGA 4211
382 rGlnValThrTrpLeuGlyThrSerLeuArgGlyGlyGlnArgTrpLeu 398
|||||
4210 AACAGGTCAAGTATCTGGGGTATCTTTAAAGAGGTCAAGATGCGCTG 4161
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
|||||
4160 ACTGAGGCCAGAAAGAGACTGTGATGGGCGACCTACTCCGAGAACCCC 4111
415 aLysGlnValArgLupheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
|||||
4110 TCGAACAACAGAGGAGTCTTCTAGGAGCGAGGCTTGTGCGCTCTGGA 4061
432 LeProGlyPheAlaThrLeuAlaAlaProLeuTrpProLeuThrLysGlu 448
|||||
4060 TCCCTGGGTTTGCAGAAATGGCAGCCCTGTACTCTCCACCAAAACG 4011
449 LysGlyGlyPheSerTrpAlaProGlnHisGlnLysAlaPheAspAlaI 465
|||||
4010 GGGACTGTGTTGAGTGGGCGCCAGACACCAAGGCCCTACCAAGACAT 3961
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
|||||
3960 CAAGCAGGCTCTTAACTGCCCTGCGCTGGGATGGCAGACTTGACTGA 3911
482 ySPProPheThrLeuTrpValAspGluArgLysGlyValAlaArgGlyVal 498
|||||
3910 AGCCCTTGAACTTTTGTGACGAGAGCAGGGCTACGCCAACAGGTGTC 3861
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaIleLysSerLys 515
|||||
3860 CTAAAGCAAAACTGCGGCTTGGCGTGGCGGCGGCTTACCTGTCGAA 3811
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
|||||
3810 AAAGCTAAACCAAGTGGCAGCTGGGCTGGCCCTTGCTACGAGTGGAG 3761
532 IaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
|||||
3760 CAGCATGCGCGTCTGACAAAGCGGTGGCAAGCTCACCATGGGACAG 3711
549 AsnIleThrValIleAlaProHisAlaLeuGlnLysIleValArgGlnPr 565
|||||
3710 CCACAGTCATCTTGGCCGCCCATGACAGTACAGCAGTAAAGCAAC 3661
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
|||||
3660 CCCTGATGCTGGGCTTCCACAGCCCAAGATGACCTACAGAGCTCTGC 3611
582 euleu...ThrGluArgValThrPheAlaProProAlaIleAlaLeuAsnPro 597
|||||
3610 TTCTGGACAGCGAGCCAGTCCAGTCCAGCAATAGTGGCCCTTAACCCA 3561
598 AlaThrLeuLeuProGlnGluThrAspGluProValThrHisAspCysH 614
|||||
3560 GCTACGCTGCTCCCT...CTACCTGAGAGAGGGCTGCACATGACTGCTCT 3514
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
|||||
3513 TGACATCTTGGCTGAAGCCACAGGAACTAGACAGATCTTACGAGCAGC 3464
631 rLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTrpVal 647
|||||
3463 CTCGCCAGAGCTGACACACCTGATACACAGTGGAGACAGCTTCTGT 3414
648 ValGluGlyLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgTh 664
|||||
3413 CAAGAGCGGAGCGCAGCGGAGCAGCAGTAAACCCAGACCGAGCGT 3364
664 rIleTrpAlaSerSerLeuProGlnGlyThrSerAlaGlnLysAlaGluL 681
|||||
3363 AGTCTGGCCAAAGCACTGCGAGCGGACATGCGCCCAAGAGCTGAGT 3314
681 eumetalLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
|||||
3313 TGATAGCGCTCACCCCAAGCCTTAAATATGSCAGAGGTAAAGAGCTGAT 3264
698 IleTrpThrAspSerArgTrpAlaPheAlaThrAlaHisValHisGlyAl 714
|||||
3263 GTTTACACCATATAGCCGTATGCTTTGCCACATGCCCATATTCACGAGA 3214
714 alleTrpLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
|||||
3213 AATATATGAAGGCGGCTTGTCTCATATGAAGGAAGAAATCAAAA 3164
731 snLysGlnGluIleLeuSerLeuLeuGlnAlaLeuHisLysProLysArg 747
|||||
3163 ATAAGCAGAGATCTTGCCCTACTGAGAGCTCTTCTCGCCCAAAAGA 3114
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleS 764
|||||
3113 CTTAGCATATTCATTGCCCCGGGACATCAGAAAGGAGCCGCGAGGC 3064
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaIleAlaVal 781
|||||
3063 AAGGGCAACAGAGATGGCCAGCCAGCGGCCGAGAAATGACCTAGA. 3015
781 alaSnLeuLeuProIleIleGluThrProLys... 791
|||||
3014 .....GAAACTCCAGAGACTTCCACACTTCTGATA 2985
792 .....AlaProGluProArgArg.....GlnTrpThrLeuGlnAs 803
|||||
2984 GAAATTCACCCCTATACTCATGAACAATTTCACTATAACGGTACTGA 2935
803 pTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGlnGlyT 820
|||||
2934 CATTAAGATCTACTAATAGTGGGCCACTTATAGCATGCAAAAG...A 2888
820 hCysTrpThrSerTrpGlyLysGluIleLeuProHisLysGlnGlyLeu 836
|||||
2887 AGGTGGGTATTCAGGGAAGCCCTGTATAGCTCATATCAATTCACCTTT 2838
837 GluTrpValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisIle 853
|||||
2837 GACATATTAATTTCTCATCAATGACCAACCTCAGTTTGTCAAAAC 2788
853 uGlnGlnLeuValArgThrSer.....ProTrpHisValLeuArgLeuP 868
|||||

```


2787 AAAGCTCTTCTAGAAAGAACTACTGCTCTTATTACATGCTGAACCGG 2738
 868 roglYValAlaAspSerValValLysHIScysValProCysGluLeuVal 884
 2737 ATCGAAGCTCAAGACATCACTGAGACTGGCCAGCCTGTCACAGGTC 2688
 885 AsnAlaAspProSerArgIleProProGlyLysArgLeuArgGlySerH 901
 2687 AATGCCAGCAAGCTGCTGCTCAACCAAGGAGACTAGGTTGAGGGCAGCG 2638
 901 sProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLys 918
 2637 ACCCGGACCCACTGGGAAATGATTCTCAAGTAAAGTAAACCTGGCCTT 2588
 918 YcIGLysAsnLysTrpLeuLeuValPheLeuAspThrPheSerGlyTrpVal 934
 2587 ATGGGATATAAATCTTTAGTTTATAGACACTTCTCTGATGGGTA 2538
 935 GluAlaLysProThrLysLysGluThrSerThrValAlaLysLysL 951
 2537 GAAGCTTCCCAACCAAGAAAGAACTGCCAAAGTTGTAACCAAGAGCT 2488
 951 eLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerA 968
 2487 ACTAGAGAAATCTCCCGAGATTGGCATGCCACAGGTAATGGACCG 2438
 968 sPasnGlyProAlaPheValAlaGluValSerGlnGlyLeuAlaLysIle 984
 2437 ACATGGGCTGCTGCTGCTGCTCAAGTAGTACAGACAGTACGCAATT 2388
 985 LeuGlyLysAspTrpLysLeuHIScysAlaLysArgProGlnSerG 1001
 2387 TTGGGCGGTGATTGAAACTACATGCTGCTTACAGACCCGAGGTTGAG 2338
 1001 YGluValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeu 1018
 2337 TCGAGTAGAAGATGATGAGCAATCAAGGAGACTTTTACATAAATTGA 2288
 1018 hrThrGluThrGlyLysAsnAspTrpMetAlaLeuLeuProPheValLeu 1034
 2287 CGCTTGAACCTGGCTAGGAGCTGGGCTGCTGCTGCTGCTGCTGCTG 2238
 1035 PheArgValAlaArgAsnThrProGlyGlnPheGlyLeuThrProArgL 1051
 2237 TATGAGCCGCGCAACAGCGCGGCGCCAGGCTCTACCCCATATGAAT 2188
 1051 uLeuThrGlyLysProProPheLeuAlaGluIleAlaPheAlaHisSerA 1068
 2187 CTATATAGGGGACACCCCGCTTGTAAACTTCCCTGATCCGACATGG 2138
 1068 LaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
 2137 CAAGGTTACTCATACCCCTCTCTCAAGCTCATTTACAGGCACTCTAC 2088
 1085 TrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaLysArgL 1101
 2087 CTGGTCCAGCAGCAAGTCTGAGACCGTGGCGGCACTTACCAAA...GA 2041
 1101 YAspLeuGln.....ValProHisArgPheGlnValGlyAspSerV 1115
 2040 GCACATACACCGGCGCGTAGTACTCAACCTTCCGGGTCGGTAGACAG 1991
 1115 aLTrpValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
 1990 TGTGGTCCGACAGACCAACCAATAAATCTAGAAACCCGCTGGAAGGA 1941
 1132 ProArgLeuValLeuLeuThrThrProThrAlaValLysValGluGlyI 1148
 1940 CCTTATACCGTCTAGTACTACCCCGCCGCTCTCAAAAGTAGAGCGAT 1891
 1148 eProThrTrpIleHisAlaSerHisValLysProAla..... 1160
 1890 TGCACGCTGATCCACGCTGCCACAGTAAAGGCTGCGACACACAGGATTG 1841

1161 ..ProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeu 1176
 1840 AGCCACCGACGAGAAATCGACATGGCGTGTTCACCGCTCCCAAAATCCCTA 1791
 1177 LysLeuArgLeuHisArg 1182
 1790 AAGATTAACATGACCCCGC 1773

seq_name: /cgn2.6/prodata/2/ina/ECTUS_COMB.seq:PCT-US93-08041-9

seq_documentation_block:

Sequence 9, Application PC/TUS9308041
 GENERAL INFORMATION:

APPLICANT: The Public Health Research Institute of the City of
 APPLICANT: New York, Inc.
 TITLE OF INVENTION: FUSION GLYCOPROTEINS
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Davis Hoxie Faithful and Hapgood
 STREET: 45 Rockefeller Pl.
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/08041
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/938,100
 FILING DATE: 28-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/_____
 FILING DATE: 20-AUG-1993
 ATTORNEY/AGENT INFORMATION:

NAME: Jacobs, Seth H.
 REGISTRATION NUMBER: 32,140
 REFERENCE/DOCKET NUMBER: 11698A50
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-757-2200
 TELEFAX: 212-586-1461

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
 LENGTH: 10367 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 PCT-US93-08041-9

alignment_scores:

Quality: 3980.00 Length: 1206
 Ratio: 3.992 Gaps: 12
 Percent Similarity: 82.670 Percent Identity: 62.687

alignment_block:

US-09-171-553b-5 x PCT-US93-08041-9/rev ..
 Align seg 1/1 to reverse of: PCT-US93-08041-9 from: 1 to: 10367

1 GYArArgGlySerAspProLeuProGluProArgValThrLeuLysVal 17
 5360 GGAGCTCAGGCTCAGAGACCCCGCCCTGAAACCCAGATACCTCAGAGT 5311


```

? Patent No.6165715
? GENERAL INFORMATION:
? APPLICANT: Collins, Mary KL
? APPLICANT: Weiss, Robin A
? APPLICANT: Takeuchi, Yasuhiro
? APPLICANT: Cossel, Francois-Lolc
? TITLE OF INVENTION: Expression systems
? FILE REFERENCE: 09/011,745
? CURRENT APPLICATION NUMBER: US/09/011,745
? EARLIER FILING DATE: 1998-06-22
? EARLIER APPLICATION NUMBER: PCT/G96/02061
? EARLIER FILING DATE: 1996-08-23
? EARLIER APPLICATION NUMBER: GB951263.1
? NUMBER OF SEQ. ID NOS: 29
? SOFTWARE: patentIn Ver. 2.0
? SEQ ID NO 3
? LENGTH: 7308
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-3

```

alignment_scores:	
Quality:	3967.00
Ratio:	3.959
Percent Similarity:	82.810
alignment_block:	
Length:	1210
Gaps:	13
Percent Identity:	62.479

US-09-1/1-553B-5 x US-09-011-745-3 ..
Align seg 1/1 to: US-09-011-745-3 from: 1 to: 7308

1 GtYArgrtGlySerAspProLeuProGluProAlaValThrLeuLysIa 17
 3184 GGAGGTCAGGGTCAGAGAGCCCGCCCTGAAACCCAGATTAACCTTAAGT 3233
 17 lGluGlyGlnProValGluPheLeuValAspThrGluIaGluHisSer 34
 3234 CGGGGGGCAACCCGTCACCTCTCGTGAATACCTGGGGCCCAACACTCCG 3283
 34 alleuLeuGlnProLeuGlyLysLeuLysGlnLysLysSerThrValMet 50
 3284 TGGTGACCCCAAAATCTGTGACCCCTGAAGGATGAATGCTGCGGGGTCAA 3333
 51 GlYAlaThrGlyGlnArgGlnProThrProThrThrArgAlaThrValAs 67
 3334 GGGGCTACGAGAGAAAGGGTATGCTGGACCAAGATGCTCAAAAGTAA 3383
 67 pleuGlyValGlyAlaArgValThrHisSerPheLeuValLeuProGluCysP 84
 3384 TCTAGCTACCGGTAAAGTACCCACTCTTCTCTCATGTACCAACACTGTC 3433
 84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnLe 100
 3434 CCTATCTCTGTGTGAAGAAGATGTTCGACTAAACATAAAGCCCAATC 3483
 101 SerPheGlu...GlnGlyArgProGluIaSerValAsnAsnLysProIle 116
 3484 CACTTTAGAGGATCAGGAGCTCAGGTTATGGGACCAATGGGGGACCCCT 3533
 116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
 3534 GCAAAGTGTGACCCCTAAATATAGAAGATGAGCATCGGCTACATGAGCCT 3583
 133 lnaValLysProAspGlnAspIleGlnSer...ThrLeuGlnGlnPhePro 148
 3584 CAAAAGACCCAGATGTTCTCTCAGGCTCCACATGCTGCTGATTTTTCCT 3633
 149 GlnAlaTrrPalaGluThrAlaGlyMetCylLeuAlaLysGlnValProP 165

165 oqlnValIleGlneUlysAlaSerAlaThrProValSerValArgLnt 182
182 yProLeuSerArgGluAlaArgGlnGlyIleThrProHisValGlnArg 198
3734 ACCCCATGTCCAAAGAAAGCCAGACTGGGGTCCAAAGCCCAATPACAGAA 3783
199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProThrAspThrPr 215
3784 CTGTGGGACCGAGGAATATAGTATACCTGGACGATGCCCTCCGTGAAACGCC 3833
215 oLeuLeuProValArgLysProGlyThrAsnAspThrArgProValGlnA 232
3834 CCTGCTACCCGCTTAAAGAAACCAAGGAGCTAATATATAGCCCTGTCACAG 3883
232 sPLeuArgGluValAsnLysArgValGlnAspLHisProThrValPro 248
3884 ATCTGAGAGAAAGTCAACAACGGGGGTGAAGACATCCACCCACAGTGGCC 3933
249 AsnProThrLysLeuLeuSerAlaLeuProProGluArgAsnThrPyThr 265
3934 AACCTTACAAACCTCTTGAGCGGGCTCCACCGCTCCACACAGTGTACAC 3983
265 rValLeuAspLeuLysAlaAlaPhePheCysLeuArgLHisProThrS 282
3984 TGTGCTTGATTAAAGAGATGCTTTTTCCTGCTGAGATCCACCCACCA 4033
282 ergLInProLeuPheAlaPheGluThrParAspProGlyThrGlyArgThr 298
4034 GTACAGCTCTCTTCGCTTGTGATGAGAAAGATCCAGAAATGGAAATCTCA 4083
299 G1yGlnLeuThrThrThrArgLeuProGlnGlyPheLysAsnSerProth 315
4084 GGCACATGTACCTCGGACCCAGACTCCCAAGGGTTTCAAAAAGTCCAC 4133
315 r1LeuPheAspGluAlaLeuHisAspArgAspLeuAlaAsnPheArgL 332
4134 CCTGTGTGATGAGGCACTGCACAGAGACTTACGACTTCGGATCCAGC 4183
332 iAspGlnValHisLeuLeuGlnThrLysAspLeuLeuAlaGly 348
4184 ACCCAGACTGTGCTCGCTACAGTACGTGATGATCTTACGTGCGCCGCC 4233
349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuL 365
4234 ACTTGTGACGTAGACTGTGCACAAAGATACGTGGGCCCTGTTCMAAACCT 4283
365 uSerAspLeuGlyThrArgAlaSerAlaLysLysValaGlnL1CysArgA 382
4284 AAGGAACTCGGGATATGGGGCTTCGGCCAAAGAAAGCCCAAAATTGGCAGA 4333
382 rGgluValThrTyLeuGlnGlyzSerLeuArgLysGlnArgThrLeu 398
4334 AACAGGTCAAGATCTGGGGTATCTTCTAAAGAGGGTCAGAGATGGCTG 4383
399 ThrGluAlaArgLysLysThrValValGlnL1LeuAlaProThrThrAl 415
4384 ACTGAGGCCAAAGAAAGAGACTGTGATGGGGCAAGCCCTACCCGAAAGCCCC 4433
415 aLysGlnValArgLysPheLeuGlyThrHisArgLysPheCysArgLeuP 432
4434 TCGACAACATAGGAGAGTTCCTAGGGAGCGCAGAGCTTCTGTGACCTTGG 4483
432 lLeuProGlyPheAlaThrLeuAlaAlaProLeuThrProLeuThrLysGlu 448
4484 TCCCTGGGTGTGGCAAAATGGCAGCCCTTGTACCCCTTCACCAAAAGC 4533
449 LysGlyLysPheSerThrAlaProGlnHisGlnLysAlaPheAspAla 465

```

4534 GGGACTGTGTTAAATTGGGCCAGACCACAAAAGGCTATCAAGAAAT 4583
455 elylsyalaleuensearialproalaleualaleuoproaspvalthr 482
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4584 CAAGCAACCTCTTAACTGCCAGCCCTGGGCTGGCGAGATTGGACTA 4633
482 yspropheThrleuThyValAspGluArgLysGlyValAlaArgGlyVal 498
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4634 AGCCCTTGAACCTCTTGTGCGAGCAAGCAAGGCTACGCCAAAGTGTTC 4683
499 leuthrGlnThrLeuGlyProThrParArgProValAlaTyrlleuSerLys 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4684 CTAAAGCAAAAACCTGGGACCTTGGCGGCGGCGGCTTACCTCCAA 4733
515 slYsleuaspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4734 AAAGCTAGACCCAGTAGACAGCTGGGCTGGCCCTTGGCTACGGAGTGTAG 4783
532 laAlaValAlaIleleuValLysAspAlaAspLysLeuThrLeuGlyGln 548
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4784 CAGCCATTGCGTACTGACAAAGGATGCGAGCAAGCTAACCATGGGACAG 4833
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPr 565
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4834 CCACTAGTCATTCTGGCCGCCCATGCGATGAGAGCACTAGTCAACAAACC 4883
565 opTroAspArgTrpMetThrAsnAlaArgMetThrHisTyrlleuSerLeu 582
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4884 CCCCAGCCCTGGCTTCCACAGCCCGGATGACTACATCAGGCTTCC 4933
582 euleu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4934 TTTTGGACACGGAGCCGGTCCAGTTGCGACCGTGGTACCCCTGAACCCG 4983
598 AlaThrleuLeuProGluGluThrAspGluProValThrHisAspCysHis 614
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4984 GCTACCGCTGCTCCA...CTGCTGAGGAAGGCGCTGCACACACAGCTGCT 5030
614 sglleuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIlePr 631
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5031 TGATATCTGTCGCGAAGCCAGACCCGACCCGACCTAACGCGACGAC 5080
631 roleuThrGlyGluValleuThrThrPheThrAspGlySerSerTyrlVal 647
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5081 CGTCCCGACAGCCGACACACACCTGTGTACCGATGGAAGACAGCTCTTA 5130
648 ValGluGlyLysArgMetAlaGlyAlaAlaValAlaAspLysThrArgThr 664
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5131 CAAGAGGAGACAGCTAAGGGGAGCGTGGTGACACCGACCGACGAGCT 5180
664 rIleThrAlaSerSerleuProGluGlyThrSerAlaGlnLysAlaGluL 681
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5181 AATGTGGCTTAAAGCCCTGCCACCGCGGACATCCGCTACGCGGCTGAAC 5230
681 eumetalaleuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5231 TGATAGACACTACCCAGGCGCTTAAAGTGGCAGAGTAAGTAAGTAAT 5280
698 IleTyThrAspSerArgTyrlAlaPheAlaThrAlaHisValHisGlyAl 714
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5281 GTTTAATCTGATAGCCGTTATGCTTTGCTTACTGCCATATCCATGAGGA 5330
714 aIleTyLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5331 AATATACAGAAAGCGCTGGTGTCTCATCTCAAGAAAGCAAGATTCAAA 5380
731 snLysGluGlnIleLeuSerleuLeuGluAlaLeuHisLeuProLysArg 747
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5381 ATAAAGACGAGATCTGGCCCTACTAAAGCCCTCTTCTGCCAAAAGA 5430
748 leuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSe 764
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5431 CTTAGCATATATCATTTGTCCAGGACATCAAAAGGACACACAGCGCGGAGGC 5480
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaVal 781
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5481 TAGAGCGAACCGGATGGCTGACCAAGCGCGCCAAAAGGACGCC..... 5523
781 alAsnleuLeuProIleIleGluThrProLysAla..... 792
5524 .....ATCACAGAGACTCCAGACACCTCTTACCTCCATATCA 5559
793 .....ProGluProArgArgGlnTyrlThrleuGluAs 803
5560 GAAATTCATCACCTACCTACCTACCTAGAACATTTTCATTACACAGTGTAG 5609
803 pTrpGlnGluIleLysLysIleAsnGln...PheSerGlnThrProGlu 819
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5610 TATAAAGCACTTAACCAAGTTGGGGCCATTATGATATAAACAAGAAAGT 5659
819 LyThrCysTyrlThrSerTyrlLysGluIleLeuProHisLysGluGly 835
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5660 ATTGGGTCTAC.....CAAGAAACCTGGATGCTGCACACAGTTTACT 5703
836 leuGluTyrlValGlnGlnIleHisArgLeuThrHisleuGlyThrLysHis 852
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5704 TTTGAATATATAGACTTTCTTCATCAGCTGACTGACTGACTGACTGCTGAAA 5753
852 slleuGlnGlnleuValArgThr.....SerProTyrlHisValleuArgL 867
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5754 AATGAAAGCTCTCTTACAGAGAACCCACAGCTTACTACATGCTGTAAC 5803
867 euproGlyValAlaAspSerValLysHisCysValProCysGlnleu 883
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5804 GGGATCGAACACTCAAAATATCTACTGAGCTGACCTCAAAAGCTTGTGCACA 5853
884 ValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgLys 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5854 GTCAACGCGCAGAGCTGCGCTTAACAGAGGAGTACAGTCCGGGGCA 5903
900 rHisProGlyAlaHisIstrGluValAspPheThrGluValLysProAla 917
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5904 TCGGCGCGGACACTATGGAGATGATTCACGAGATTAAGCCCGGAT 5953
917 ySTyrlGlyAsnLysTyrlleuLeuValPheValAspThrGlyPheSerGlyTrp 933
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5954 TGTATGCTATTAATATCTCTAGTTTATATGATACCTTTCTTGCTGG 6003
934 ValGluAlaTyrlProThrLysLysGluThrSerThrValAlaLysLys 950
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6004 ATAGAACCTTCCCAACCAAGAAAGAACCCGACAGTGTAAACCAAGA 6053
950 slleuLeuGluGluIlePheProArgPheGlyIleProLysValIleGlyL 967
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6054 GCTACTAGAGAGATCTTCCAGGTTCGCGATGCTCAGGATATGGGAA 6103
967 eTrAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuLys 983
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6104 CTGACATATGGCTGCTTCTGTCGTAAGGTAGTACAGACAGTGGCCGAT 6153
984 IleLeuGlyIleAspTrpLysLeuHisCysAlaTyrlArgProGlnSerSe 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6154 CTGTTGGGATGATGGAATTTACATGTGATACAGACCCCAAGGCTC 6203
1000 rGlyGlnValGluArgMetAsnArgThrIleLysGluThrleuThrLysL 1017
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6204 AGCCAGGTAAAGAAAGTAATGAAGCATCAAGGAGACTTTAACTAAAT 6253
1017 euthrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6254 TAAAGCTTGAACCTGCTGCTAGAGATGGGTGCTCTACTCCCTTAGCC 6303
1034 leuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrl 1050
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6304 CTGTACGAGCCCGCAACACAGCGCGGCCCATGAGCTTACCCCATATGA 6353

```


4034 GTCAGCTCTCTCCCTTTGAGTGGAGAGATCCAGAGATGGAAATCTCA 4083
299 GlyLeuThrLeuThrArgLeuProGlnGlyPheLysAsnSerProth 315
|||||
4084 GGACAAATTGACCTGGACGAGACTCCACAGGGTTTCAAAAACACTCCAC 4133
315 rLePheAspGlnAlaLeuHisArgAspLeuAlaAsnPheArgIleGln 332
|||||
4134 CCGTTTATGATGAGCATGACACAGAGACTGACAGACTTCGGATCCAGC 4183
332 isProGlnValThrLeuLeuGlnThrValAspAspLeuLeuAlaGly 348
|||||
4184 ACCGAGACTGATCTCTGCTACAGTACGTGATGACTTACTCTGCTGGCCGC 4233
349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuGlnLe 365
|||||
4234 ACTCTGAGCTATGACTGCCAACAGGTACTCGGGCCCTGTTACAAACCT 4283
365 uSerAspLeuGlyTyrArgAlaSerAlaLysAlaGlnIleCysArg 382
|||||
4284 AGGGAACCTCGGGATCGGGCTCGGCCAAGAAAGCCAAATTTGCCAGA 4333
382 rGlnValThrLysLeuGlyTyrSerLeuArgGlyGlnArgTrpLeu 398
|||||
4334 AACAGCTCAAGTATCTGGGGTATCTTTAAAGAGGTACAGAGATGGCTG 4383
399 ThrGlnAlaArgLysLysThrValAlaGlnIleProAlaProThrAl 415
|||||
4384 ACTGAGGCCAGAAAGAGACTGTGATGGGGCAGCTACTCCGAGAACCCC 4433
415 aLysGlnValArgLysPheLeuGlyThrAlaGlyPheCysArgLeuTrp 432
|||||
4434 TCGACAACTAAAGGAGTCTTAGGAGCGCAGGCTTGTGCGCTCTGTA 4483
432 leProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGln 448
|||||
4484 TCCCTGGTTCGACAAATGACGACGCCCTTGTACCTCTCTCAACAAAG 4533
449 LysGlyGlyPheSerTrpAlaProGlnHisGlnLysAlaPheAspAla 465
|||||
4534 GGGACTCTGTTTAATTGGGGCCAGACCAACAAAGGCTATCAAGAAAT 4583
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThr 482
|||||
4584 CAAGCAAGCTCTTAACTGACCCCAAGCCCTGGGGTTGCCAGATTGGACA 4633
482 ySerProPheThrLeuThrValAspGlnArgLysGlyValAlaArgGlyVal 498
|||||
4634 AGCCCTTGAACCTTTGTGCGAGAGAGACAGGCTACCCCAAGGTGTC 4683
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLys 515
|||||
4684 CTACCCCAAAACTGGGACCTTGGCGTGGCGGCTGCTACCTGCTCAAA 4733
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIle 532
|||||
4734 AAAGCTAGACCCAGTAGCAGCTGGGTGGCCCTTGGCTAGCGAGTGGTG 4783
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGln 548
|||||
4784 CAGCCATTGCGCTGACTGACAAAGATGCGAGCAAGCTAACATGGGACAG 4833
549 AsnIleThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnPr 565
|||||
4834 CCACTAGTCACTTGGCCCCCCTATGAGAGGCTAGTCAAAACACACC 4883
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeu 582
|||||
4884 CCCCAGCCGCTGCTTCCAAAGCCCGGAGTACTCACTATCAGGCTTGC 4933
582 eLeu...ThrGlnArgValThrPheAlaProProAlaAlaLeuAsnPro 597
|||||
4934 TTTTGGACAGAGACGGGCTCCAGTTGAGACCGGTGGTAGCCCTTGAAACCG 4983
598 AlaThrLeuLeuProGlnGlnThrAspGlnProValThrHisAspCysHis 614
|||||
4984 GCTACGCTGTCCCA...CTGCCGTAGAGAAAGGCTCCAAACACATGCGCT 5030
614 sGlnLeuLeuIleGlnGlnThrGlyValArgLysAspLeuThrAspIle 631
|||||
5031 TGATATCTGTGGCCGAAGCCCGGAAACCCGACCCGACTTAAAGGACACAG 5080
631 rLeuThrGlyGlnValLeuThrTrpPheThrAspLysSerTrpVal 647
|||||
5081 CGCTCCCAAGGCCGACCCAGACCTGTGATCAGGATGGAAGACTCTCTTA 5130
648 ValGlnLysArgMetAlaGlnAlaValAlaValAspLysArgThr 664
|||||
5131 CAGAGGAGACAGCGTAAAGCGGAGAGCTGGGTGATCCACCGAGACCGAGT 5180
664 rIleTrpAlaSerSerLeuProGlnGlyThrSerAlaGlnLysAlaGln 681
|||||
5181 AATCTGGCTAAAGCCCTCGCAGCGCGGACATCCGTCAGCGGGCTGAAAC 5230
681 eMetAlaLeuThrGlnAlaLeuArgLeuAlaGlnGlyLysSerIleAsn 697
|||||
5231 TGATAGCAGTCCAGCCAGGCTTAAAGATGGCAGAGAGTAAAGAGCTAAAT 5280
698 rIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
|||||
5281 GTTATATCTGATACCGCTTATGCTTGTGCTAGTCCCATTCATGATGAGAG 5330
714 aIleTyrLysGlnArgLysLeuLeuThrSerAlaGlyArgGlnLysAl 731
|||||
5331 AATATACAGAAAGCGCTGGTGTCTCATCATCAGAGGAAAGAGATCAAAA 5380
731 sLysGlnGlnIleLeuSerLeuLeuGlnAlaLeuHisLeuProLysArg 747
|||||
5381 ATAAAGCAGAGATCTGGCCCTACTAAAGCCCTTTGTGCCCCAAAGA 5430
748 LeuAlaIleIleHisCysProGlnHisGlnLysAlaLysAspLeuIle 764
|||||
5431 CTTAGCATATATCCATTGTGCCAGACATCAAAAGGACACAGCGCGAGGC 5480
764 rArgLysAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAla 781
|||||
5481 TAGAGGCAACCGGATGGCTGACCAAGCGGCCGCAAGGACGAC... 5523
781 aLysLeuLeuProIleIleGlnThrProLysAla... 792
5524ATCACAGAGACTCCAGACACTCTACCTCTCATATA 5559
793ProGlnProArgArgGlnThrThrLeuGlnAs 803
5560 GAAATATGATACCTACCTACCTCAGAACATTTTCATATACAGAGACTGA 5609
803 rTrpGlnGlnIleLysLysIleAspGln...PheSerGlnThrProGln 819
|||||
5610 TATAAGGACCTTAACCAAGCTTGGGGCCATTATGATTAACCAAGAGAGT 5659
819 LysThrCysTyrThrSerTrpGlyLysGlnIleLeuProHisLysGlnGly 835
|||||
5660 ATTGGGCTTAC...CAAGGAAACCTGATGATCCGAGCAGATTAAT 5703
836 LeuGlnThrValGlnGlnIleHisArgLeuThrHisLysGlnLysHis 852
|||||
5704 TTTGAAATTATTAAGACTTCTTCATCAGCTGATGATCAGCTCAGCTCTCA 5753
852 sLeuGlnGlnLeuValArgThr...SerProTyrHisValLeuArgL 867
|||||
5754 AATGAAGGCTCTCCTAGAGAGAGAGCCAGACGCTCCCTACATGCTGAAC 5803
867 eUrProGlyValAlaAspSerValValLysHisCysValProCysGlnLeu 883
|||||
5804 GGGATCGAACACTCAAAATATACCTGAGACCTGCAAAAGCTTGTGACAA 5853


```

884 ValAspAlaAsnProSerArgIleProProGlyLysArgLeuValGlySe 900
|||||..... ||| ... ||| ..... |||
5854 GTCAACGCGCAAGCTGCGCTGCTAAACAGCACTAGGCTCCGCGGCA 5903
900 rHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaL 917
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
5904 TGGGCGCGGCACTCATGTGGAGATGATTCACCCAGATAAAGCCGGAT 5953
917 ySTyGlyAsnLysTyrrLeuLeuValPheValAspThrPheSerGlyTrp 933
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
5954 TGTATGGCTATAAATATCTCTAGTTTATATAGATACCTTTTCGGCTGG 6003
934 ValGluAlaTyrrProThrLysGlySerThrValValAlaLysLys 950
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6004 ATAGAAAGCTTCCCAACCAAGAAAGAACCCCAAGGTCGTAACCAAGAA 6053
950 sIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlyS 967
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6054 GCTACTAGAGAGATCTCCCGAGGTGGCATGCTCAGTATTTGGGAA 6103
967 eRAspAsnGlyProAlaPheValAlaGluValSerGlnGlyLeuAlaLys 983
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6104 CTGAAATAGGCGCTGCTCTCCAGGTAGTACAGACAGTGGCCGAT 6153
984 IleLeuGlyIleAspTrpLysLeuHisCysAlaTyrrArgProGlnSerSe 1000
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6154 CTGTGGGCGATTTGATGAAATACATTTGCTATACAGACCCCAACCTC 6203
1000 rGlyGluValGluArgMetAsnArgThrIleLysGluThrLeuThrLysL 1017
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6204 AGGCCAGGTAGAAAGATGATAGACCATCAAGAGACCTTAACTAAAT 6253
1017 eutThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6254 TAAAGCTTGCAACGTGCTAGAGACTGGGTCTCTACTACCTCCCTTAGCC 6303
1034 LeuPheArgValAlaArgAsnThrProGlyGlnPheGlyLeuThrProTyrgl 1050
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6304 CTGTACCGAGCGCCCAACAGCGCGGCGCCCATGCGCTCACCCCATATGA 6353
1050 uLeuLeuTyrglyGlyProProProLeuAlaGluIleAlaPheAlaHis 1067
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6354 GATCTTATATGGGCGCACCCGCCCTTGAACCTTCCTGACCCGACCA 6403
1067 eRAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1083
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6404 TGACAAAGATTACTACAGCCCTCTCTCAAGCTCACTTACAGGCTCTC 6453
1084 GluThrValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTrpSerG 1100
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6454 TACTTAGTCCAGCAGAAAGTGGAGACCTCTGGCGGCGACCTTACCA.. 6501
1100 yGlyAspLeuGln.....ValProHisArgPheGlnValGlyAspS 1114
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6502 .GAACAACCTGGACCGGCGGTGCTACCTCACCTTACCGAGTGGGACAA 6550
1114 eRValTyrrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLys 1130
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6551 CAGTGTGGGTCCGCGACACCAAGCTAAGAACTAGAACCTCGCTGAAA 6600
1131 GlyProTyrrLeuValLeuLeuThrThrProThrAlaValIleValGluG 1147
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6601 GAGACTTACAGACTCTCTGCGACACCCCGCGCTCAAAAGTAAACGG 6650
1147 yIleProThrTrpIleHisAlaSerHisValLysProAlaProPro.... 1162
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6651 CATCGCAGCTTGATACAGCGCCCAAGTGAAGGCTGCGACCCCGGGG 6700
1163 .....ProspSer.....GlyTrpArgAlaGluLysThrLysLysPro 1175
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6701 GTGGACCATCTCTTACAGTACGACATGCGCGCTTCAACGCTCTCAAAA 6750
1176 LeuLysLeuArgLeuHisArgLeuValPro 1185

```

```

|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
6751 TTAATAATAGGTAAACCGCGAGGCCCC 6780

```

```

seq_name: /cgn2-6/ptodata/2/1na/6B.COMB.seq:US-09-011-745-2

```

```

seq_documentation_block:
; sequence 2, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 7616
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-2

```

```

alignment_scores:
Quality: 3967.00 Length: 1210
Ratio: 3.959 Gaps: 13
Percent Similarity: 82.810 Percent Identity: 62.479

```

```

alignment_block:
US-09-171-553b-5 x US-09-011-745-2 ..

```

```

Align seg 1/1 to: US-09-011-745-2 from: 1 to: 7616

```

```

1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysVal 17
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
2345 GAGAGTCAGAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 2394
17 lGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerV 34
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
2395 CGGGGGGCAACCGCTACCTTCTGCTGATAGTGGGGGCCCAACACTCCG 2444
34 allLeuLeuGlnProLeuGlyLysLeuLysGlyLysLysSerTrpValMet 50
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
2445 TCCTACACCAAAATCTCGAGCCCTTAAGTGAATAGTGTGCTGGGTCCAA 2494
51 GluAlaThrGlyGlnArgGlnTyrrProThrThrArgArgTrpValAs 67
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
2495 GGGGCTACTGAGGAAAGCGGATGCGTCGACACAGGATGCCAAAGTACA 2544
67 pLeuGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
2545 TCTACTACCGGTAAAGTCACCCACTCTTCTCCATGATACAGACGTGTC 2594
84 rovalProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
2595 CCTATCTCTTTTGAAGAGATTTGCTGACTAACTAAAGCAAAATC 2644
101 SerPheGlu...GlnGlyArgProGluValSerValAsnLysProIle 116
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||
2645 CACTTGAAGGATCAGAGATCAGTATATGGACCAATGGGCGAGCCCT 2694
116 eThrValLeuThrLeuGlnLeuAspAspGluTyrrArgLeuTyrrSerPro 133
|||||..... ||| ..... ||| ..... ||| ..... ||| ..... |||

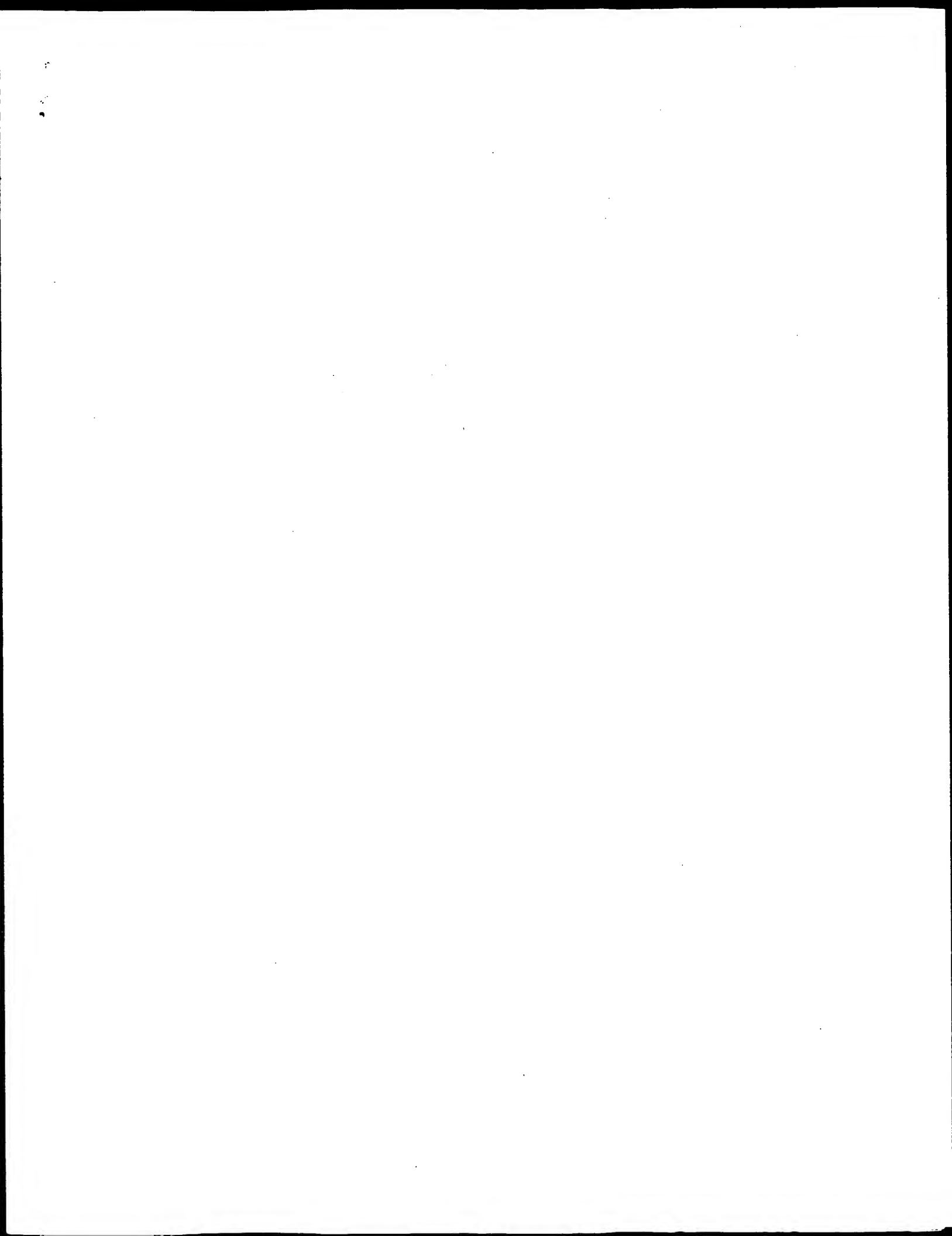
```

```
2695 GCAAGTGTGACCCCTAAATATAGAAAGATGAGACTCGGCTACATGAGACCT 2744
133 InValIleuProAspGlnAspIleuGlnSer...TyrLeuGlnGlnPhePro 148
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2745 CAAAGAGCCAGATGTTCTCTAGGGTCCACATGGCTGTGATTTCT 2794
149 GlnAlaThrPalaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPr 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2795 CAGGCTTGCGGGAACCGGGGCGATGGGACTGGGACGTTCCGCAAGCTCC 2844
165 oGlnValIleuGlnIleuLysAlaSerAlaThrProValSerValArgGlnT 182
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2845 TGTGATCATACCTCTGAAGCAACCTTACCCCGTGCTCCATAAACAAT 2894
182 TyrProLeuSerArgGlnAlaArgGlnGlyIleTyrProHisValGlnArg 198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2895 ACCCATGTCACAGAAACCCAGACTGGGAGTCAAGCCCACTACAGAGA 2944
199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTyrAsnThrPr 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2945 CTGTGGACACGAGGAACTGTGTAACCTGACCGACGTCCTGGAACAGCC 2994
215 OleuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnA 232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2995 CTGCTACCCGTTAAGAAACGAGGACTAATGATTATAGGCTGTCCAG 3044
232 SPLeuArgGlnValAsnLysArgValGlnAspIleHisProThrValPro 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3045 ATCTGAGAGAACTCACAGAGCGGTGGAGAGACATCCACCCACGCTGCC 3094
249 AsnProTyrAsnLeuLeuSerAlaLeuProProGlnArgAsnTyrTyrTh 265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3095 AACCTTACACACTTGTGAGCGGCTCCACGCTCCACGAGTGTGTACAC 3144
265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3145 TGTCTGTGATTTAAAGATGCTTTTCTGTGCTGAGACTCCACCCACCA 3194
282 eGlnProLeuPheAlaPheGlnTyrPArgAspProGlyThrGlyArgThr 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3195 GTCACGCTCTTTCGCCCTTGTAGTGGAGAGATCCAGAGATGGGAATCTCA 3244
299 GlnGlnLeuThrTyrThrArgLeuProGlnGlyPheLysAsnSerProTh 315
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3245 GGACAATTGACCTGGACACGACTCCACAGGGTTCAAAAAACAGTCCAC 3294
315 rIlePheAspGlnAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3295 CCTGTTTGATGAGGACTGCAAGAGACTAGCAGACTTCCGGATCCAGC 3344
332 ISProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGly 348
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3345 ACCGAGACTTATCTGCTACAGTACGTGATGACTTACTGCTGGCCGCC 3394
349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuGlnLeu 365
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3395 ACTTCTGAGTACGACTGCCAACAGAGTACTCGGCTGTACAAACCT 3444
365 userAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3445 AGGGAACCTCGGATCGGGCTCGGCCAAAGAACCCCAATTGGCCAGA 3494
382 rGlnGlnValThrTyrLeuGlyTyrSerLeuArgGlyGlnArgTyrLeu 398
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3495 AACAGGTCAAGTATCTGGGGTATCTTCTAAAGAGGGTCAAGATGGCTG 3544
399 ThrGlnAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3545 ACTGGGCGCAGAAAGAGACTGTATGGGCGACCTACTCCGAGAACCCC 3594
415 alyGlnValArgLysPheLeuGlyThrAlaGlyPheCysArgLeuTyrP 432
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3595 TCGACAACCTAAGGAGTTCCTAGGGACGCGAGGCTTCTGTGCTTGTGA 3644
432 IeProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGln 448
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3645 TCCCTGGTGTTCAGAAATGACAGCCCTGTACCTCTGCACCAAAAGC 3694
449 LysGlyGlyPheSerTyrPalaProGlnHisGlnLysAlaPheAspAlaI 465
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3695 GGGACTCTGTTTAATTGGGGCCGAGACCAAAAGAGCCCTATCCAGAAAT 3744
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3745 CAAGCAAGCTCTTCTAATGCTCCAGCCCGGGGTTGCCAGATTTCAGTA 3794
482 YSPProPheThrLeuTyrValAspGlnArgLysGlyValAlaArgGlyVal 498
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3795 AGCCCTTTGAGACTTGTGTGACGAGAGAGAGGCTACCGCAAAAGTGTTC 3844
499 LeuThrGlnThrLeuGlyProTyrPArgArgProValAlaTyrLeuSerLy 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3845 CTAAACCAAAACTGGGACTTGGCTGGCGCGGCTGACCTGTCCAA 3894
515 sLysLeuAspProValAlaSerGlyTyrProValLysLeuLysAlaIleA 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3895 AAAGCTAGACCCAGTAGACAGCTGGTGCCCTTCTCTACGAGTGTAG 3944
532 IaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3945 CAGCCATGCGCTACTGACAAAGATGCAAGCAGCTAACATCAGGAGAG 3994
549 AsnThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnPr 565
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3995 CCAGTATGTCATCTGGCCCCCATATGCAGTAGAGGCACTAGTCAAAACAC 4044
565 oProAspArgTyrMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4045 CCCCAGCCGCTGGCTTCCAAACCCGAGATGATCTCTACAGGCTTCC 4094
582 eLeuLeu...ThrGlnArgValThrPheAlaProProAlaIleLeuAsnPro 597
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4095 TTTTGGACAGCGACCGGCTCAGTGTGAGACCGGTGTACCCGTAACCCG 4144
598 AlaThrLeuLeuProGlnGluThrAspGluProValThrHisAspCysH 614
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4145 GCTACCGCTGCTCCA...CTGCCGTGAGAAAGGCTCAACACAACTGCT 4191
614 sGlnLeuLeuIleGlnGluThrGlyValArgLysAspLeuThrAspIleP 631
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4192 TGAATATCTGGCCGAAGCCGACCGGAACCCGACCTAACGAGACCA 4241
631 rOleuThrGlyGlnValLeuThrTyrPheThrAspLysSerTyrVal 647
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4242 CGCTCCAGAGCGCCGACACACCTGCTACAGGATGGAAAGCACTCTTA 4291
648 ValGlnGlyLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgTh 664
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4292 CAAGAGGAGCAGCGCTAAGCGGGAGCTCGGATGACCACCGAGAGGT 4341
664 rIleThrPalaSerSerLeuProGlnGlyThrSerAlaGlnLysAlaGlnL 681
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4342 AATCTGGCTAAAGCCCTCCAGCCGAGCATCTCGCTCAGCGGGCTGAAC 4391
681 eumAlaLeuThrPheGlnAlaLeuArgAlaGlnLysSerIleAsn 697
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4392 TGATACAGCTCAACCCAGCGCTTAAATATGCGAAGAGTAAAGCTAAT 4441
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4442 GTTATATCATGATAGCCGTTATGCTTTGCTACTGCCATATCATAGGAA 4491
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGlyIleLysA 731
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4492 AATATACAGAAAGGCGTGGTGTGCTACATCAGAAAGCAAGAGATCAAAA 4541
```

[illegible]

Mon Feb 25 07:44:24 2002

us-09-171-553b-5.feb22std.rni



OM of: US-09-171-553b-6 to: Issued_Patents_NA:* out_format : pfs

Date: Feb 24, 2002 10:18 AM

About: Results were produced by the GenCode software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09171553/runat_22022002_122826_11171/app-query.fasta.1.3312
-db=Issued Patents_NA -QFMT=fastap -SUFFIX=feb22std.rni
-GAPOP=12.000 -GAPOP=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-LOOPEXT=0.000 -GAPOP=4.500 -GAPOP=0.050 -GAPOP=10.000
-XGAPEXT=0.500 -GAPOP=6.000 -GAPOP=7.000 -XGAPEXT=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blotsum62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORX=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000
-USPR=US09171553 @CGN1.158 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-171-553b-6
Query length: 656
Database: Issued Patents_NA*
Database sequences: 351203
Database length: 113238999
Search time (sec): 134.760000

Score_list:	Strid	Orig	Zscore	Escore	len	Documentation	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-111-085-3			3433.00	5859.62	0.0	3482	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-111-085-1			2694.50	4421.86	1.2e-238	2462	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-766-528-2			2497.50	4204.08	1.4e-228	7333	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-766-528-1			2360.00	4001.38	2.4e-215	8060	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-766-528-3			2353.00	3991.22	1.2e-214	8132	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-376-781-25			2136.00	3635.91	7.1e-195	2000	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-376-781-30			2136.00	3635.91	7.1e-195	2000	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-376-781-24			1627.00	2764.19	2.6e-146	1493	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-111-745-7			1508.50	2541.95	6.1e-134	6312	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-716-351A-1			1486.50	2500.26	1.3e-131	8535	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-716-351A-2			1431.00	2404.62	2.7e-126	8655	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-110-300A-8			1299.00	2178.09	1.1e-113	8323	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-866-642-8			1299.00	2178.09	1.1e-113	8323	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-011-745-6			1258.00	2111.62	5.7e-110	6061	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-850-961-1			1258.00	2107.55	9.7e-110	8332	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-850-961-2			1235.50	2087.65	1.2e-108	1914	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-011-745-5			1231.50	2066.11	2.0e-107	6028	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-850-961-13			1227.00	2070.97	8.7e-108	2001	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-258-420-9			1226.00	2070.97	1.1e-107	1965	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-258-420-13			1208.00	2021.76	5.8e-105	8202	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-258-420-7			1204.50	2021.15	3.8e-105	8325	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-011-745-9			1191.50	2011.99	2.0e-104	1911	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-552-369-19			1173.50	1980.58	1.1e-102	1979	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-105-483-310			1163.50	1960.40	1.5e-101	2499	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-105-483-310			1163.50	1960.40	1.5e-101	2499	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-458-101-310			1163.50	1960.40	1.5e-101	2499	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-116-351A-5			1024.50	1702.43	3.6e-87	10970	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-376-781-18			640.00	1076.30	2.7e-52	698	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-073-836-1			634.00	1040.55	2.6e-50	5109	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-073-836-2			634.00	1040.55	2.6e-50	5176	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-278-630A-7			626.50	1044.12	1.7e-50	1407	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-484-126-2			626.50	1043.77	1.7e-50	1446	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-278-630A-8			601.50	1008.73	1.5e-48	1329	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-278-630A-9			601.00	1000.57	4.4e-48	1374	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-484-126-4			600.00	999.02	5.4e-48	1356	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-376-781-10			484.00	819.75	5.2e-38	278	
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-376-781-4			433.00	685.87	1.5e-33	10306	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-376-781-8			430.00	726.96	7.7e-33	276	
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-011-745-1			423.50	687.53	1.2e-30	2518	

/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-011-745-8 + 423.50 676.73 4.8e-30 5665
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-007-282B-1 + 423.00 691.66 7.1e-31 170
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-716-351A-3 + 420.50 665.19 2.1e-29 966
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-376-781-7 + 396.00 662.38 3.0e-29 445
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-691-563C-58 + 295.50 472.23 1.2e-18 17

seq_name: /cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-111-085-3

seq_documentation_block:

; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4236/75168
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GH 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

alignment_scores:

Quality: 3433.00 Length: 658
Ratio: 5.257 Gaps: 2
Percent Similarity: 99.240 Percent Identity: 98.176

alignment_block:

US-09-171-553b-6 x US-09-111-085-3

Align seg 1/1 to: US-09-111-085-3 from: 1 to: 3482

1 MethisProthleuSerArgRHisLeuProthraRgylgylupr 17
|||||
911 ATGATCCACGTTAAAGTGGCGCCACCTCCGAGCTGGGCGAGAGCC 960
17 OlyArgLeuArgRgylleProLeuSerPheAlaSerIleAlaTrpPheLeu 34
|||||
961 GAAAGATCGAGATCCCTTAACCTCGCTCCATCGCTGCTTCA 1010
34 hrlleuThrlleThrProGlnAlaSerLysArgLeuIleAspSer 50
|||||
1011 CTTAAACATTAACCTCCAGCCAGTAAACCTTATAGACAGCTCG 1060
51 AsnProHlsArgProLeuSerLeuThrlleuLeuIleAspProAsp 67
|||||
1061 AACCCCATTAACCTTATCCCTTAACCTGCTGATTAAGCCTGATAC 1110
67 rglValThrlleThrValAsnSerLysArgValAlaProArgLysTrp 84
|||||
1111 GGGTCTCACTGAATATACACACGAGGTGCTCTAGAGGACCTG 1160
84 rPrProGlnLeuHisPheCysLeuArgLeuIleAspProAlaValLys 100
|||||
1161 GCCCTGAACCTGATTTGCTCCATTAATTAACCCGCTGTAAGAC 1210
101 ThrProAsnLeuValArgSerLysLysPheThrlleThrCysProGly 117
|||||
1211 ACACCTCCACCTAGTCCAGTATGATTAATGCTCCAGGAC 1260
117 rglValThrlleThrValAsnSerLysArgValAlaProArgLysTrp 134
|||||
1261 AGAGAAAGAAATTAATCTGGGGGCTTCTGCGGATCTTCTGAGGAT 1310
134 rPserCysValThrlleSerAsnAspGlyAspTrpLysTrpProLysSer 150

1311	GGAGCTGGGTCACCTCCCAAGATGGAGACTGGAAATGGCCGATCTCTCTC	1360
151	GLnspyrtyValLysPheSerPheValAsnSerGlyProGlyLysTyrLy	167
1361	CAGAGCCGGGTAATAATCTCTTTCATATCCGGCCGGGCAAGTAAAG	1410
167	MetMetLysLeuTyrLysAspLysSerGlySerProSerAspLeuAspT	184
1411	AGTATGAAACTATATTAAGATTAAGAGCTGCCTCCATCAGCTTAAGTT	1460
184	tyrLeuLysIleSerPheThrGlu. ArgLysThrGlyLysTyrSerLysV	200
1461	ATCTAAGAGATTAAGTTTCACATGAAAAGAAAACAGAAATATTTCAAAG	1510
200	alaAspLysTrpTyrGluLeuGlyAsnSerPheLeu. LeuTyrGlyGlyG	216
1511	TGGATTAATATGGTATGAGCTGGGGAAATAGTTTTTTTAAATATGGCGGGG	1560
216	yalagLysThrLeuThrIleArgLeuAsnArgIleGluThrGlyThrGluP	233
1561	AGCAGGGTCCACTTATTAACATTTGGCTTATGAGTAAAGAGGGGACAGAC	1610
233	roProValAlaMetGlyProAspLysValLeuAlaGluGlyProPro	249
1611	CCCTGTGGCAGTGGGACCCGATTAAGTACTGGCTGAACAGGGGCCCCG	1660
250	AlaLeuGluProProHisAsnLeuProValProGluLeuThrSerLeuAr	266
1661	GCCTGTGAGCACCCGATTAATCTGGCGGGGCCCAATTAACCTGCTGGC	1710
266	gProAspIleThrGlnProProSerAsnSerThrThrGlyLeuIleProT	283
1711	GCCTGACATTAACACGCCCTCAGAACGGGACACATGAGTATGATTCCTA	1760
283	hrAsnThrProArgAsnSerProGlyValProValLysThrGlyGlnArg	299
1761	CCAACACGGCTTGAAATCTCCAGAGTGTCTCTTTAGACAGGACAGAGA	1810
300	LeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspR	316
1811	CTCTTCAGTCTATCCAGGAGACTTTCCAACGCATCAACTGCACCGACC	1860
316	oAspAlaThrSerSerGlyTripleCysLeuSerSerGlyProProTyrT	333
1861	TGATGCGCACTTCTTCTGTGGCTTGTCTATCCCTCAGGGGCTCCTAAT	1910
333	tyrGluGlyMetAlaLysGluArgLysPheAsnValIleThrLysGluHisArg	349
1911	ATGAGGGGATGGCTTAAGACAGAAATATTCATGTGACCCAAAGACATAGA	1960
350	AsnGluCysThrTyrGlySerArgAsnLysLeuThrLeuThrGluValSe	366
1961	AATCATATGATACATGGGGGTCCCGAAATAAGTTCACCTCAGTAAGTTTC	2010
366	rgLysLysGlyThrCysIleGlyLysAlaProProSerHisGlnHisLeuc	383
2011	CGGGAGAGGGGAGATGCAATGAGAAAGCTCCCATCCACCAACACCTTT	2060
383	ysTyrSerThrValValTyrGluGlnIleSerGluAsnGlnTyrLeuVal	399
2061	GCTATAGTACTGTGTTTATATGACAGGCCCTCAGAAATACAGTATTAAGTA	2110
400	ProGlyLysAsnArgTyrTrpAlaCysAsnThrGlyLeuThrProCysVa	416
2111	CTGTGTTAATACAGGTGTGGTGCATGCAATACGTGGGTATACCCCCGTGT	2160
416	IserThrSerValPheAsnGlnSerLysAspPheCysValMetValGlnI	433
2161	TTTCACCTCAGTCTTCAACCATCCAAAGATTTCTGTGTCATGTGTCCAA	2210
433	IleValProArgValTyrThrHisProGluGluValValLeuAspGlyTyr	449

```

2211 TCGTCCCGCCAGGTGTAAGTACTACCACTCTGAGAGAAAGTGTCTCTTATGTAATAT 2260
450 ASPIRYTATGTyrAsnArgProLysAlaGluProValSerLeuThrLeuAl 466
|||||
2261 GACATATGGTATAAACGACCAAAAAGAAAGAACCCGTATCTCCCTTACCTGAGC 2310
466 aValMetLeuGlyLeuGlyThrAlaValAlaGlyThrGlyThrAla 483
|||||
2311 TGTAATCTCTGGATTAGGACGCGCCCTTGCGCTAGGAAACAGGACGACTG 2360
483 lalauilletHrGlyProGlnGlnLeuGluLysGlyLeuGlyLeuLeu 499
|||||
2361 CCTGATACAGAGACCAACAGACAGCTAGAGAAAAGACTGGTGTAGCTACAT 2410
500 AlAlaMetHrGluAspLeuAlaValAlaLeuLysGluSerValSerAsnLe 516
|||||
2411 GCGGCGCATGACAGAAAGATCTCCGAGCCTTAGAGGAGTCTGTTAGCAACT 2460
516 uGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArg 533
|||||
2461 AGAAGAGTCCCTGACTTCTTGTGTGAAGTGGTTCTTACAGAACCGGAGG 2510
533 lYleuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLys 549
|||||
2511 GATTAGTCTGCGTGTCTTAAGAGAAAGGGGTTATGTGCAGCCTTAA 2560
550 GluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMe 566
2561 GAAGAAATGTTGCTTCTATGTAGTACATCACAGAGGCATCACAGACTCAT 2610
566 tAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluAla 583
|||||
2611 GACCAAGACTTAAGAGAAAGGTTAGAGAGCGCTGGAAGGGAAGAAAGAGCTG 2660
583 spGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTyrMetHrThr 599
2661 ACCAGGGGGTGTGAAGATGTCTTCAACAGAGTCTCTCTTGATATCACACC 2710
600 LeuLeuSerAlaLeuHrGlyProLeuValValLeuLeuLeuLeuLeu 616
|||||
2711 CTGCTTTCTGCTCTGACGGAGCCCTTAGTACGCTGCTGTTACTTAC 2760
616 rValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgV 633
|||||
2761 AGTGTGGCCCTGCTTAATTATAGATTGTGCTTGTGTAGAGAACGAG 2810
633 alSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeuLeu 649
|||||
2811 TGAAGTCACTCCAGATCATGTACTTAGGCACACACTACCAAGGCGCTTCTG 2860
650 SerGlnGlyGluThrAspLeu 656
|||||
2861 AGCCAAGAGAGAAACTGACCTC 2881

seq_name: /cogr2_6/ptodata/2/ina/6A_COMB.seq:US-09-111-085-1
seq_documentation_block:
: Sequence 1, Application US/0911085
: Patent No. 6100034
: GENERAL INFORMATION:
: APPLICANT: Stoye, Jonathan P
: APPLICANT: Weiss, Robin A
: TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
: FILE REFERENCE: 4238/75168
: CURRENT APPLICATION NUMBER: US/09/111,085
: EARLIER APPLICATION NUMBER: 1998-07-07
: EARLIER FILING DATE: 1997-05-16
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 2462

```


TYPE: DNA
 ORGANISM: Porcine retrovirus
 US-09-111-085-1

alignment_scores:
 Quality: 2594.50 Length: 680
 Ratio: 4.420 Gaps: 11
 Percent Similarity: 86.324 Percent Identity: 73.088

alignment_block:
 US-09-171-553b-6 x US-09-111-085-1

Align seg 1/1 to: US-09-111-085-1 from: 1 to: 2462

```

1 MethSprothLeuSerArgArgHisLeuProthArgGlyGlyLeu 17
  |||||||
211 ATGATCCACGCTAAGCGCGCCACCTCCCGATTCCGGGGTGAAGCC 260
  |||||||
17 OlysrArgLeuArgLeuProLeuSerPheAlaSerIleAlaTrpPheLeu 34
  |||||||
261 GAAAGACGAGAAATCCCTTAAGCTTCGCTCCATCGCGTGTCTTA 310
  |||||||
34 hLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSer 50
  |||||||
311 CTCTGTCAATTAATCTCTCAAGTTAATGGTAAAGCGCTGTGTGACAGCCG 360
  |||||||
51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAspTh 67
  |||||||
361 AACTCCCAATAAACCTTATCTCTACCTGTACTACTGACGACGCGTAC 410
  |||||||
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
  |||||||
411 AGGTATTAATTAATTAACGACCTCAAGGAGGCGCTCCCTGGGACCTG 460
  |||||||
84 rProGlnLeuHisPheCysLeuArgLeuIleAsnProAlaVal 98
  |||||||
461 GGGCTGAATTAATATGTCCTCGATCGATCAATCCCTGGTCAATGAC 510
  |||||||
99 LysSerThrProAsnLeuValArgSerTrpGlyPheTrpCysCysPr 115
  |||||||
511 CAGCCACACACCCCGATGATCTACTCCGTGCTACGGGTTTACGTTTCCC 560
  |||||||
115 ogly...ThrGluGluLysTrpCysGlySerGlyGlySerPheC 131
  |||||||
561 AGGACCCCAATTAATGAAGAAATATTGTGAATCCTCAGATTTCTTT 610
  |||||||
131 ysrArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
  |||||||
611 GCAAGCAATGAGCTCATTAATCTTCTTAATGATGGGAATGGAAATGGCA 660
  |||||||
148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSer 161
  |||||||
661 GTCTCTCAGCAGACAGAGTAATCTTCTTTGTTAACAATCCTACAG 710
  |||||||
162 .....GlyProGlyLysTrpLys.....MetMetL 170
  |||||||
711 TTATTAATCAATTAATTAATGCGCATGGAGATGGAAGAAATGGCAACAC 760
  |||||||
170 ysrLeuTrpLysAspLys.....SerCysSerProSerAsp 181
  |||||||
761 GGGTACCAAAAGATGACGAATTAAGCAAAATTAAGCTGTCACTTGTTAGAC 810
  |||||||
182 LeuAspTrpLeuLysIleSerPheThrGlu..ArgLysThrGlyLysTrp 197
  |||||||
811 CTAACTACTTAATAAATAGTTTCACTGAGAAAGAAAGAAACAGAAATAT 860
  |||||||
198 SerLysValAspLysTrpTrpGluLeuGlyAsnSer..PheLeuLeuTrp 214
  |||||||
861 TCAAAAGTGGTAATGTATATCTTGGGGAATAGTGTAC.....TATG 904
  |||||||
214 LysGlyGlyAla.....GlySerThrLeuThrIleArgLeuArgIle 227
  |||||||

```

```

905 GAGGCTCTGGAGAGAAAGAAAGATCTGTCTGACTATTGCGCTCAGAAATA 954
228 GluThrGlyThrGluProProValAlaMetGlyProAspLysValLeuAl 244
  |||||||
955 GAAACTCAGATGGAACCTCCGTTGCTATAGACCAAAATTAAGGTTGGC 1004
244 aGluGlnGlyProProAlaLeuGluProProHisAsnLeuProValProG 261
  |||||||
1005 CGAACAAGGACCTCCAAATCCCAAGACAG..... 1032
261 hLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSerThr 277
  |||||||
1033 .....AGGCATCTCTTAACCCCTCTGTGATTAACATTAACAC 1068
278 ThrGlyLeuIleProThrAsnThrProArgAsnSerProGlyValProva 294
  |||||||
1069 TCTGGATCAATGCCCT.....GAGCTTAACATCACTAT 1103
294 hLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAla 311
  |||||||
1104 TAAACACGGCGGCAAACTTTTACCTCATCCAGGAGGCTTTTCAACCTC 1153
311 LeuSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSer 327
  |||||||
1154 TTAACTCCAGACCTCCAGAGGCTACTCTTCTTGTGGCTTGTAGCT 1203
328 SerGlyProProTrpTrpGlyGlyMetAlaLysGluLysLysPheAsnVa 344
  |||||||
1204 TCGGCGCCACCTTACTATAGAGGAATGCTAGACGAGGAAATTCAAATGT 1253
344 hThrLysGlnHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeu 361
  |||||||
1254 GACAAAGACATATGAGACCAATATACATGGGATGCCAAATTAAGCTTA 1303
361 hLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProPro 377
  |||||||
1304 CCCTTACTGAGGTTCTGGAAAGGACCTGCATAGGATGATGCTCCCA 1353
378 SerHisGlnHisLeuCysTrpSerThrValValGlyGlnAlaSerG 394
  |||||||
1354 TCCACACACACCTTTGTAACACACCTGAACCTTTATTAACACCTGTA 1403
394 uAsnGlnTrpLeuValProGlyTrpAsnArgTrpTrpAlaCysAsnThrG 411
  |||||||
1404 GAGTAAATATCTGTAACCTGTTATGACAGAGTGGGCAATGATACTG 1453
411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPhe 427
  |||||||
1454 GATTAAACCCCTGTGTTCACCTGTGTTCACCAACCAACTTAAGACTTT 1503
428 CysValMetValGlnIleValProArgValTrpTrpHisProGluGluVa 444
  |||||||
1504 TGGCTTAATGTCCAATTTGCCCGGCTGTACTACTATCCCGAAAAAGC 1553
444 lValLeuAspGluTrpAspTrpArgTrpAsnArgProLysArgLysPro 461
  |||||||
1554 AGTCTTGATGATATGACTATAGATTAATGCGCAAAAGAGAGCCCA 1603
461 aLysLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyAl 477
  |||||||
1604 TATCCCTGACACTAGCTGTAAATGCTCGATGGATGGGCTGCGACAGCGTG 1653
478 GlyThrGlyThrAlaAlaLeuThrGlyProGlnGlnLeuGluLysG 494
  |||||||
1654 GGAACGAGAGAGGCTGCCCTAATCAAGAGACCGCAACACTGAGAAAGG 1703
494 yLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 511
  |||||||
1704 ACTTAATACCTAATCAATGCAATTTGTAAAGGAAGATGCCAAGCCCTA 1753
511 lSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValAla 527
  |||||||
1754 AATCTGTCAAGTAACTGAGGAATCCCTTAACCTCTTATCTGAAGTGT 1803

```

```

528 LeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlyGlyLeu 544
1804 CTRACAGACAGAGAGGGGTTAGATCTGTTATTTCTAAAGAGAGGGTT 1853
544 UCysAlaAlaLeuLysGluGlyCysCysPheTyrValAspHisSerGly 561
1854 ATGTGAGCTTAAAGAGAGATGCTGCTTATATGATCACTCAGAGAG 1903
561 IallArgAspSerMetLysLeuArgLysLysLeuGluLysArgArg 577
1904 CCATCAGAGACTCCATGAGCAAGCTTAGAAGAAAGTTAGAGAGCGCTGA 1953
578 ArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPheAsnArg 594
1954 AGGGAAGAGAGCTGACACAGGGGTGTTGAGAGATGTTCAACAGCTC 2003
594 rProThrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValAla 611
2004 TCCTTGATGACACACCTGCTTCTGCTCTGACGGGGCCCTAGTAGTCC 2053
611 LeuLeuLeuLeuThrValGlyProCysLeuLeuAsnArgPheValAla 627
2054 TGCTCCTGTTACTTACAGTGGCTTGTCTTAATTAAGTTGTTGCC 2103
628 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGln 644
2104 TTGTGTAGAGACAGAGTAGTGCAGTCCAGATCATGTACTTAGGACACA 2153
644 nTyrGlnGlyLeuLeuSerGlnGlyGluThrAspLeu 656
2154 GTACCAAGGCTCTGAGCCACAGAGAACTGACCTC 2190
seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:US-08-766-528-2
seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MCP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-2
;
alignment_scores:
    Quality: 2497.50      Length: 678
    Ratio: 4.328          Gaps: 13
    Percent Similarity: 85.103      Percent Identity: 71.829
alignment_block:
    US-09-171-553B-6 x US-08-766-528-2
Align seg 1/1 to: US-08-766-528-2 from: 1 to: 7333
4 ThrLeuSerArgArgHisLeuProThrArgGlyGlyLeuProLysArgLeu 20
4757 ACCTTACCTGCTACTTTTGACACACACAGCGGCTGTGAAAGTCGA.... 4801
20 uArg.IleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThr 36
4802 .AGGAATCCCTTAAGCTTCGCTCCATCGCGGTCTCTACTCTGTCA 4850
37 IleThrProGlnAlaSerSerLysArgLeuLeuAspSerSerAsnProH 53
4851 ATAACTCTCAAGTAAATGTAAGCTTGTGAGACAGCCGCAACTCCCA 4900
53 sArgProLeuSerLeuThrTrpLeuIleLeuAspProAspThrGlyAlT 70
4901 TTAACCTTATCTCTCCTGCTGTTACTTACTGACCTCCGCTACAGATTA 4950
70 hrValAsnSerThrArgGlyValAlaProArgGlyThrTrpProGlu 86
4951 ATATTACAGACACTCAGAGGAGGCTCCCTTGGGACCTGGTGCTGAA 5000
87 LeuHisPheCysLeuArgLeuIleAsnProAlaVal.....LysSerTh 101
5001 TTATATGTCCTGCTTCAGTACAGTAATCCCTGCTCAATGACACAGCCAC 5050
101 rProProAsnLeuValArgSerTyrGlyPheTyrCysProGly....T 117
5051 ACCCCCGATGTAGTCGCTGCTTACGGGTTTACCTTCCGACGACCCC 5100
117 hgGluLysGluLysTyrCysGlySerGlyGlySerGlySerPheCysArg 133
5101 CAATTAATGAAGAATATTGTGGAATCCTCAGGATTTCTTTTGAAGCAA 5150
134 TrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 150
5151 TGGACTCTCAATAACTTAAATGATGCGAATGGAATGCCAGCTCTCA 5200
150 uGlnAspArgValLysPheSerPheValAsnSer..... 161
5201 GCAACACAGAGTAAGTACTCTTTTGTTAACAATCTCCACGATTATATC 5250
162 .....GlyProGlyLysTyrLys.....MetMetLysLeuTyr 172
5251 AATTTAATTATGAGCATGAGATGAGAAAGTTGCAACAGCGGTTCAA 5300
173 LysAspLys.....SerCysSerProSerAspLeuAspTyr 184
5301 AAAGTGTGACAAATTAAGCAAAATAGCTGTCATTGCTTAACCTAGATT 5350
184 rLeuLysIleSerPheThrGlu...ArgLysThrGlyLysTyrSerLysVa 200
5351 CTTAAATAATAGTTTCACTGAAAGAGAAACAGAGAAATATTCATAAAGT 5400
200 LAspLysTrpTyrGluLeuGlyAsnSer..PheLeuLeuTyrGlyGly 216
5401 GGTAAATGTAATCTTGTGGGAATAGTATAC.....TATGAGGCTCT 5444
217 Ala.....GlySerThrLeuThrIleArgLeuArgIleGluThrG 230

```


TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-766-528-1

alignment_scores:
Quality: 2360.00 Length: 656
Ratio: 4.252 Gaps: 11
Percent Similarity: 84.604 Percent Identity: 68.445

alignment_block:
US-09-171-553b-6 x US-08-766-528-1

Align seg 1/1 to: US-08-766-528-1 from: 1 to: 8060

```

1 MethSProThrLeuSerArgAlrghisLeuProThrArgglyGlyLeu 17
  |||||
86 ATGCATCCACGTTAAGCGCGCACCTCCGATTCGGGGTGAAGCC 135
  |||||
17 olysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeu 34
  |||||
136 GAAAGACTGAAATGCCCTTAAGCTTCGCCCTCCATCGGTGTCCTTA 185
  |||||
34 hrLeuThrIleThrProGlnAlaSerSerIysArgLeuIleAspSer 50
  |||||
186 CTCTGCATATACCTCTCAGACTATGATGCCATAGACACGCCCTG 235
  |||||
51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAsp 67
  |||||
236 AACGCCATAAACCTTATCTCTCAGCTGTTAATTACTGACTCGGCAC 285
  |||||
67 rglValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
  |||||
286 ACGTTATATATCAACAACACTCAAGGAGGCTCTTTAGGAACCTGT 335
  |||||
84 rProGlnLeuHisPheCysLeuArgLeuIleAsnProAlaValIysSer 100
  |||||
336 GGCCTGATCTATACCTTGGCCACATGATTTCTAGTCTACCTCA 385
  |||||
101 ThrProAsnLeuValArgSerTrpGlyPheTrpCysCysProGlyTh 117
  |||||
386 ...CCCCAGATATCTCCATGCTCAGCGATTATGTTGCCAGAC 432
  |||||
117 rGluysGlu...LysTrpCysGlyGlySerGlyGluSerPheCysArg 133
  |||||
433 ACCAAATATGGAACATGTGGGAATCCAGAGATTCTTTGTAAAC 482
  |||||
133 rglTrpSerCysValThrSerAsnAspGlyAspTrpIleAspProIle 149
  |||||
483 AATGGAACCTGTATACCTCTATATGATGATTTGGAATGGCCAACTCT 532
  |||||
150 LeuGlnAspArgValLysPheSerPheValAsn.....SerGlyProG 164
  |||||
533 CAGCGAGATAGGTAAGTTTCTTATGTCAACACCTATACCACTCGG 582
  |||||
164 YLSTYrLysMetLysLeuTrpLysAspLysSer.....CysSerP 179
  |||||
583 ACAATTTATTAACCTGACCTGGATTAGAACTGGAAGCCCAAGGCTCTC 632
  |||||
179 rOserAspLeuAspTrpLeuLysIleSerPheThr..GluArgLysThr 195
  |||||
633 CTTCGACCTAGATTACCTAAATAAGTTTCACTGAGAAAGGAAAAACA 682
  |||||
195 lYlYsTrpSerLysValAspLysTrpTrpGluLeuGlyAsnSer..Phe 211
  |||||
683 GAAATATATCTAAATAGGTAAATGATGTCTGGGGAATGATATAT.. 730
  |||||
211 uLeuTrpGlyGlyAla.....GlySerThrLeuThrIleArgL 225
  |||||
731 ...TATGAGGCTCGGGTAAACAACACGCTCATCTTAACATATTCGCC 776
  |||||
225 euArgIleGluThrGlyThrGluProValAlaMetGlyProAspLys 241
  |||||
777 TCAAAATTAACAG...CTGGAGCTCCCAATGCTATAGACCAAAATACG 823
  |||||
242 ValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 258
  |||||
824 GTCTTGAAGGGTCAAGACCCCAACCAAGGACCA..... 859
  |||||
258 oValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSera 275
  |||||
860 .....GACACATCTCTA 872
  |||||
275 snSerThrThrGlyLeuIleProThrAsnThrProAlaGlnSerProGly 291
  |||||
873 ACATTAACCTTGTGATCAACCCCACTGATGCTACACAGACACT..... 916
  |||||
292 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnAlaPh 308
  |||||
917 .....AAATGGGGCAAACTTTTATAGCTCATCCAGGAGGCTTT 957
  |||||
308 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeu 325
  |||||
958 TCAAGCTCTTAACCTCAGACGCTCAGAGCTACTCTTCTTGTGGCTAT 1007
  |||||
325 YsLysSerSerGlyProProThrTrpGlnGlyMetAlaLysGluArgLys 341
  |||||
1008 GCTTAGCTTTGGGCCACCTTACTATGAAAGAAATGCTTAAGAGGAAA 1057
  |||||
342 PheAsnValThrLysGlnHisArgAsnGlnCysThrTrpGlySerArg 358
  |||||
1058 TTCAATGTGCAAAAGAACATAGACACCATGACATGGGATCCCAAAA 1107
  |||||
358 nLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLys 375
  |||||
1108 TAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1157
  |||||
375 lAProProSerHisGlnHisLeuCysTrpSerThrValValTrpGluG 391
  |||||
1158 TTCCCCCATCCCAACACCTTGTATACCACTGAAGCTTAAATCA 1207
  |||||
392 AlAserGluAsnGlnTrpLeuValProGlyTrpAsnArgTrpPheLacy 408
  |||||
1208 ACCTGTAAAGTCAATATCTGTACCTGTTATGACAGGCGTGGGCGAG 1257
  |||||
408 sAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSer 425
  |||||
1258 TAACTAGCATTAACCCCTGTGTTCACCTTGTTTTAAACCAAACTA 1307
  |||||
425 YsAspPheCysValMetValGlnIleValProArgValTrpTrpHisPro 441
  |||||
1308 AAGATTTTTCATTAATGATGTCCAATATGTTCCCGAGTATATGATGCC 1357
  |||||
442 GluGluValValLeuAspGluTrpAspTrpArgTrpAsnArgProLysAr 458
  |||||
1358 GAAAGAACATCTCTGATGATATGACTACAGAAATCATCGACAAAGAG 1407
  |||||
458 gGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaV 475
  |||||
1408 AGAACCATATCTCTGACACTTCTGTGATGCTGAGCTTGGAGTGGCAG 1457
  |||||
475 aLysValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 491
  |||||
1458 CAGGTGTAGACAGACAGACAGCTGCTGCTGTCAGGACACAGACGCTA 1507
  |||||
492 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAl 508
  |||||
1508 GAACAGAGACTTAGTAACCTACATGCAATTTGAACGAAGATCTCCAAAC 1557
  |||||
508 aLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerG 525
  |||||
1558 CCTAGAAAATATCTGTACGTAACCTGAGGAATCCCTTAACCTCTATCTG 1607
  |||||
525 lValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlu 541
  |||||
1608 AAGTATCTCTACAGATTAAGAGGGTTAGATTATTTATTTCTAAAGAA 1657

```

```

542 GLYGLILEUCYSALAALALEULYGLUCYSCYPHETRYVALASP 558
|||||
1658 GGAGATTATGTAGACCTTGAAGAGAGATGCTTTATATGATGATCA 1707
|||||
558 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeu 575
|||||
1708 TTCAGGGGCCATCAGAGACTCCATGAACTAGAGAAAGTTGGAGA 1757
|||||
575 rGaIArgArgLysGluArgLysGlnGlyTrpPheGluGlyTrpPhe 591
|||||
1758 AGCGTCGAGGAGAAAGAACTACTGAGGCTTGAAGGATGCTTC 1807
|||||
592 AsnArgSerProTrpMetThrThrLeuSerAlaLeuThrGlyProle 608
|||||
1808 AACAGGCTCTGTTGGCTACCTTCTTCTTTACAGAGACCTT 1857
|||||
608 uValValLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgp 625
|||||
1858 AATAGTCTCTCTCGTTACTACAGATGGCCATGATATTAACAAGT 1907
|||||
625 heValAlaIleArgLysValSerAlaValGlnIleMetValLeu 641
|||||
1908 TAATGGCTTCATTAGAGAACGATTAAGTCAGATCATGCTACT 1957
|||||
642 ArgGlnGlnTrpGln 646
|||||
1958 AGACACAGTACCAA 1972
|||||
seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-766-528-3

```

```

seq documentation block:
? Sequence 3, Application US/08766528
? Patent No. 6190861
? GENERAL INFORMATION:
? APPLICANT: Jay A. Fishman
? TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
? TITLE OF INVENTION: AND METHODS OF USE
? NUMBER OF SEQUENCES: 74
? CORRESPONDENCE ADDRESS:
? ADDRESSER: LAHIVE & COCKFIELD, LLP
? STREET: 60 State Street
? CITY: Boston
? STATE: Massachusetts
? COUNTRY: USA
? ZIP: 02109-1875
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/766,528
? FILING DATE:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/572,645
? FILING DATE: 14-DEC-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Louis Myers
? REGISTRATION NUMBER: 35,965
? REFERENCE/DOCKET NUMBER: MGP-038CP
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617)227-7400
? TELEFAX: (617)227-5941
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8132 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-08-766-528-3

```

```

alignment_scores:
    Quality: 2353.00    Length: 656
    Ratio: 4.232        Gaps: 12
    Percent Similarity: 84.756    Percent Identity: 68.598

```

```

alignment_block:
US-09-171-553b-6 x US-08-766-528-3

```

```

Align seg 1/1 to: US-08-766-528-3 from: 1 to: 8132

```

```

1 MetHisProThrLeuSerArgArgHisLeuProThrArgGlyGluPr 17
5620 ATGCATGCCACGCTTAACCGGCGCCACCTCCGATTCGGGGGAGAAAGCC 5669
|||||
17 oLysArgLeuArgLleProLeuSerPheAlaSerIleAlaTrpPheLeu 34
|||||
5670 GAAAGACGTGAATCCCTTAAGCTTCGCTCCATCGGTGGTGGTCTTA 5719
|||||
34 hIleuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
|||||
5720 CTCTGTCATTAACCTCTCAGACTAATGATAGGCGATAGAGACACCTG 5769
|||||
51 AsnProHisArgProLeuSerLeuThrTrpLeuIleIleAspProAsp 67
|||||
5770 AACTCCCATTAACCTTATCTCTCACCTGGTAAATTAATTAAGTCCGCGAC 5819
|||||
67 rGlyValAlaThrValAsnSerThrArgGlyAlaAlaProArgGlyTrp 84
|||||
5820 AGGTATTAATATCAACACCTCAAGGAGGAGGCTCTTAAGAACCTGGT 5869
|||||
84 rProGlnLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
|||||
5870 GGCCTGATGTCATACGTTGGCTCAGATCAGTATTTCCTAGTCTGACCTCA 5919
|||||
101 ThrProProAsnLeuValArgSerTrpGlyPheTrpCysCysProGlyTh 117
|||||
5920 ...CCCCAGATATCCCTCATGCTCAGAGGATTTATGTTGGCCAGAGACC 5966
|||||
117 rGluLysGlu...LysTrpCysGlyGlySerGlyGlySerPheCysArg 133
|||||
5967 ACCAATTAATGAAACATTTGCGGAAATCCAGACATTTCTTTTGAAC 6016
|||||
133 rGTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProLysSer 149
|||||
6017 AATGGAACCTGTATACCTCTAATGATGATGGAATGGAATGGCCAACTCT 6066
|||||
150 LeuGlnAspArgValLysPheSerPheValAsn.....SerGlyProG 164
|||||
6067 CAGCAGGATAGGGTAAAGTTTCTTATGTCACACCTATACGACGCTCTGG 6116
|||||
164 yLysTrpLysMetIleLysLeuTrpLysAspLysSer.....CysSerp 179
|||||
6117 ACAATTTAATTAACCTGACTGACTGATTAAGACTGGAAGCCCAAGTGTCTC 6166
|||||
179 rSerAspLeuAspTrpLeuLysIleSerPheThr...GluArgLysThrG 195
|||||
6167 CTTCAGACCTAGATTACCTTAATAATAAGTTCACTGAGAAAGAAAACA 6216
|||||
195 lYsTrpSerLysValAspLysTrpTrpGluLeuGlnAsnSer.PheLe 211
|||||
6217 GAAATATCTCTAATAATGGTAAATGTATGCTTGGGGAATGGTAAAT.. 6264
|||||
211 uLeuTrpGlyGlyGlyValA.....GlySerThrLeuThrIleArgL 225
|||||
6265 ....TATGAGGCTCGGGTAAACACACAGGCTCATATTACTATATTCGCC 6310
|||||
225 euArgLleGluTrpGlyThrGluProProValAlaMetGlyProAspLys 241
|||||
6311 TCAAAATTAACCAAG...CTGGAGCTTCATATGCTATAGACCAATAATAG 6357
|||||
242 ValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 258
|||||

```

6358	GTCTTGACGGGTCAAGAACCCCAACCCAAAGGACA.....	6393
258	oValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA	275
6394	6406
275	snSerThrThrGlyLeuIleProThrAsnThrProArgsnSerProGly	291
6407	ACATAATCTTGATGCATGACAGCCCACTGAGTCT.....AACAGC.....	6444
292	ValProValIleThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaP	308
6445	...ACGACATAAATGGGGGCAAAACATTTTATGCTCATGACAGAGCTTT	6491
308	eglnAlaIleAsnSerThrAspProAspAlaThrSerGlyStrIleuG	325
6492	TCAAGGCTCTTAATCTCCAGCATCTCCAGAGGCTACCTCTTCTGTGGCTAT	6541
325	ysLeuSerSerGlyProProTyrTyrGlnGlyMetAlaLysGlnArgLys	341
6542	GCTTACTCTGGGGCCCACTTACTATGAAAGATGGCTTAGAAGAGGAA	6591
342	PheAsnValThrLysGlnIleArgAsnGlnCysThrThrPglIserArgAs	358
6592	TTCAATAGTGACAAAGAAGACATAGAGCAACATGCACATGGGGATCCCAAA	6641
358	lysLeuLeuThrLeuThrGlnValSerGlyLysGlyThrCysIleGlyLysA	375
6642	TAAAGCTTACCCTTACTAGAGGTTTCGGAAAGAGCACTGCATAGAGAAAG	6691
375	laProProSerHisGlnHisLeuGlyCysTyrSerThrValIleTyrGlnGln	391
6692	TTCCCCCATCCCAACCAACACCTTTTAAACACACTGAAGCTTTAAATCAA	6741
392	AlaSerGlnAsnGlnTyrLeuValProGlyTyrAsnArgTyrPrtPalacy	408
6742	ACCTCGAGAGCAATATGCTGTACTGCTGATTATGACAGTGCTGGCATG	6791
408	snsnThrGlyLeuThrProGlyValSerThrSerValPheAsnGlnSerL	425
6792	TATATCGATTAATACCCCTTGTTGGTTTCCACCTTGTTTAAACCAACTA	6841
425	ysAspPheCysValMetValGlnIleValProArgValTyrTyrHisPro	441
6842	AAGATTTTGGCATTTGGTCCAAATGTGTTCCCGAGTGAATTAATCACTCC	6891
442	GlnGlnValValLeuAspGlnTyrAspTyrArgTyrAsnArgProLysAs	458
6892	GAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATGCAACAAAGAG	6941
458	gsluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaV	475
6942	ACAACCCATATCTCGACACTGTGTGATGCTCTGGACTTGGAGTGGAG	6991
475	alGlnValGlyThrGlyThrAlaIleLeuIleThrGlyProGlnGlnLeu	491
6992	CAGGCTGAGGACAGAGAAAGACTGCTCCGTGCACGGGACCAAGACACTA	7041
492	GlnLysGlyLeuGlyGlyLeuHisAlaAlaMetThrGlnAspLeuArgL	508
7042	GAAACAGAGCACTTAGAACCTGCATCGAATTTGTAAACAGAAATCTCCAAAG	7091
508	aleuLysGlnSerValSerAsnLeuGlnGlySerLeuThrSerLeuSerG	525
7092	CCTAGAAAAATCTGTGAGTAACTCGAGAGAACTCCCTAACTCCTTATCTG	7141
525	luValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGln	541
7142	AAGTAGTCTCTACAAATAAGAAAGGGTTAATATTAATTTCTAAAGAA	7191
542	GlyGlyLeuCysAlaAlaLeuLysGlnGlyCysCysPheTyrValAspH	558
7192	GGAGATTAATTTGTAGACCTTGAGAGAGCAATGCTGTTTATATGTGATCA	7241

558 sserGlyAlaIleArgAspSerMetAspLysLeuArgLysLysLeuGUA 575
|||||
7242 TTCAGGGCCATCAACAGACTCCATACACAGCTTAGAATAAGGTGGGAGA 72919
575 rGArGArGArgLuarGluAlaaspGInGlyTPRPheGluGlyTPRphe 591
|||||
7292 AACGTCGAAGGAAAGGAACCTACTCAGAGGTGGTTTGAGGCAATGCTTC 73419
592 AsnArGserProTPrMetThrThrLeuLeuSerAlaLeuThrGlyProLe 608
|||||
7342 AACAGGTCCTCTGGTGTGGTACCTACTTCTGCTTTAACAGGACCTT 73919
608 uValValLeuLeuLeuLeuLeuThrThValGlyProCysLeuIleAsnArgP 625
|||||
7392 AAATGTCCTCCCTCCGTACTACACAGTTGGGCGCATGATATTATTAACACAGT 74419
625 heValAlaPheValArgGluArgValSerAlaValGlnIleMetValLeu 641
|||||
7442 TTAATGCTTCAATTAGAACAGAAATAGTGCAGTCCACATATGCTACTT 74919
642 ArgGInGInTyTyrGln 646
|||||
7492 AGACAACAGTACCA 7506

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-376-781-25

```
seq_documentation_block:
```

Sequence 25, Application US/09376781
Patent No. 6361806

: GENERAL INFORMATION:

; APPLICANT: Banerjee, Papia T.

APPLICANT: Patience, Clive

INVENTOR: MOLECULAR
AFRICANI, ANDREISSON, GOLDI

; Patent No. 6261806

TITLE OF INVENTION: Use

FILE REFERENCE: 61/30-26/
CURRENT APPLICATION NUMBER: 11

CURRENT FILING DATE: 1999-08-

EARLIER APPLICATION NUMBER: 6

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 25

LENGTH: 2000
TYPE: DNA

ORGANISM: Artificial Sequence

; FEATURE:

OTHER INFORMATION:	Description:
OTHER INFORMATION:	Description:

OTHER INFORMATION: COMPILED
US-09-376-781-25

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

alignmen_scores: 2136 00

Ratio: 3.963

Percent Similarity: 82.416 P

Argument Block:

US-09-171-553B-6 x US-09-376-78

Align seg 1/1 to: US-09-376-78

1 Methisprothril eusera arar

43 ATACATCCCACGTTAAGCCGGCG

17 of us [Arct]e [Arct]e [Arct]e

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

93 AAAAAGACTAAAAATCCCCCTTAAG

34 `brt.auth.brt]emhrProc]a1]a`

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleaspSer 50

```
|||||..... |||.....: |||..... |||.....
143 CTCTGCAATACCTCTCAGACTAATGATATGACATAGAGACACCTTG 192
|||||..... |||.....: |||..... |||.....
51 AsnProHisArgProLeuSerLeuThrTrpLeuIleLeuAspProSph 67
|||||..... |||.....: |||..... |||.....
193 AACCCCATTAACCTTATCTCTGACCTGGTTAATTACTGACTCTGACAC 242
|||||..... |||.....: |||..... |||.....
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
|||||..... |||.....: |||..... |||.....
243 AGGATTAATATCCACAGCGCTGAGGAGGAGCTCCTTAGAAACCTGGT 292
|||||..... |||.....: |||..... |||.....
84 rProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
|||||..... |||.....: |||..... |||.....
293 GGCCTGATCTATATGCTGCTCAGATCATCTCTAGCTGACCTCA 342
|||||..... |||.....: |||..... |||.....
101 ThrProProAsnLeuValArgSerThrGlyPheThrCysCysProGlyTh 117
|||||..... |||.....: |||..... |||.....
343 ACC...CCAGATATCTCCCTGCTTACGATTTTATGTGGCCAGGACC 389
|||||..... |||.....: |||..... |||.....
117 rGluLysGlu...LysThrCysGlyLysGlySerGlyLysSerPheCysArg 133
|||||..... |||.....: |||..... |||.....
390 ACCAAATATGAAACAACTATGGAATCTAGAGATTCTTTACAAAC 439
|||||..... |||.....: |||..... |||.....
133 rGTPSerCysValThrSerAsnAspGlyAspTrpLysTrpProLysSer 149
|||||..... |||.....: |||..... |||.....
440 AATGAGAGCTGTGTAACTTATGATGAAATCGAAATGCGCAACCTCT 489
|||||..... |||.....: |||..... |||.....
150 LeuGlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTy 166
|||||..... |||.....: |||..... |||.....
490 CTGCGAGGATGAGGTAGCTTTTCTTATGCTAC...CCCATTAACCA 533
|||||..... |||.....: |||..... |||.....
166 rLysMetLysLeu...TyrLysAspLysSer...CysSerProSerA 181
|||||..... |||.....: |||..... |||.....
534 CCGGACCTGAAACAGTCATACAGATACCATTTCTGGTTCCTCCAG 583
|||||..... |||.....: |||..... |||.....
181 sPLeuAspTyrLeuLysIleSerPheThrGlu...ArgLysThrGlyLys 197
|||||..... |||.....: |||..... |||.....
584 ACCTAGATATCTTAAATAAGTTTACCGCAAAAAAACAAGAAATAT 633
|||||..... |||.....: |||..... |||.....
197 rYSerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeuLeuTyr 213
|||||..... |||.....: |||..... |||.....
634 ATCCAAATATGATTAATGATATGCTCGGGGAAATAT...ATATTATAC 680
|||||..... |||.....: |||..... |||.....
214...GlyGlyGly.....AlaGlySerThrLeuThrIleArgLeuArgIleG 228
|||||..... |||.....: |||..... |||.....
681 AGGTGGGACAGACAAACGAGCTCATCTTAACCATCGACTTAATAA 729
|||||..... |||.....: |||..... |||.....
228 LuThrGlyThrGluProProValAlaMetGlyProAspLysValLeuAla 244
|||||..... |||.....: |||..... |||.....
730...AGCCAGCTAGAGCTCCCAAGGCTATAGAGCGAATAGCGTCTTAACG 777
|||||..... |||.....: |||..... |||.....
245 GluGlnGlyProProAlaLeuGluProHisAsnLeuProValProG1 261
|||||..... |||.....: |||..... |||.....
778 GGTCAAGAAGACCCCAACCCAGGACCATCTCTGATATA..... 816
|||||..... |||.....: |||..... |||.....
261 nLeuThrSerLeuArgProAspLysIleThrGluProProSerAsnSerThr 278
|||||..... |||.....: |||..... |||.....
817.....ACTTCTAAATTAGACCCCACTGAG.....TCTAACACACAGA 853
|||||..... |||.....: |||..... |||.....
278 hGlyLeuIleProThrAsnThrProArgAsnSerProGlyValProVal 294
|||||..... |||.....: |||..... |||.....
854 CT..... 855
|||||..... |||.....: |||..... |||.....
295 LysThrGlyGluArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaI 311
|||||..... |||.....: |||..... |||.....
856 AAAAGGGGACAAACTTTTATGCTCATCCAGGAGCTTTCAAGCTCC 905
|||||..... |||.....: |||..... |||.....
311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuGlyLeuSerS 328
|||||..... |||.....: |||..... |||.....
906 TAACTCCAGACTCCAGAGCTACCTCTTCTGTGGCTTGTCTTAACCT 955
|||||..... |||.....: |||..... |||.....
328 eGlyProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVal 344
|||||..... |||.....: |||..... |||.....
956 CGGGCCACACTTACTATAAGAAATGCGTAAAGAGAAAAATTCATGTG 1005
|||||..... |||.....: |||..... |||.....
345 ThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuTh 361
|||||..... |||.....: |||..... |||.....
1006 ACAAAAAAACAATAGAGACCAATGATAGGGATCCCAAAATAGCTTAC 1055
|||||..... |||.....: |||..... |||.....
361 rLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProS 378
|||||..... |||.....: |||..... |||.....
1056 CCTTACTGAGGTTTCTCGAAAGACACCTGCATTAATAAAGTTCCCCCAT 1105
|||||..... |||.....: |||..... |||.....
378 eHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGlu 394
|||||..... |||.....: |||..... |||.....
1106 CCAACCAACACTTTTACCAACCACTGAGAGCTTTAATCAACCTCTGAG 1155
|||||..... |||.....: |||..... |||.....
395 AsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrG1 411
|||||..... |||.....: |||..... |||.....
1156 AGTCATATCTGTTACTGTTATGACAGGTGGGCGATGAAATACATG 1205
|||||..... |||.....: |||..... |||.....
411 rLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheC 428
|||||..... |||.....: |||..... |||.....
1206 ATTAACCCCTTGCTGTTCCACCTTGTTTCAACCAACTAAAGACTTTT 1255
|||||..... |||.....: |||..... |||.....
428 yValMetValGlnIleValProArgValTyrTyrHisProGluGluVal 444
|||||..... |||.....: |||..... |||.....
1256 ACATATATGTCCAATTTGTTCCCGAGTATATTACTATCCAAAGAAACA 1305
|||||..... |||.....: |||..... |||.....
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProVa 461
|||||..... |||.....: |||..... |||.....
1306 ATTCTGATGATATATGATTACAGAACCATCGCAAAACAAACCACTAT 1355
|||||..... |||.....: |||..... |||.....
461 rSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValG 478
|||||..... |||.....: |||..... |||.....
1356 ATCCCTGACACTCGCAGTATGCTCGAGCTCGAGAGATTAACAGGTGA 1405
|||||..... |||.....: |||..... |||.....
478 rLysThrGlyThrAlaAlaLeuIleThrGlyProGlnLeuGluLysGly 494
|||||..... |||.....: |||..... |||.....
1406 GAACAGAACTGCAGCTTATGTTAGTACAGACCTCAGAGCTAGAAACAGA 1455
|||||..... |||.....: |||..... |||.....
495 LeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG1 511
|||||..... |||.....: |||..... |||.....
1456 CTTAGTACCTACATCAATTTGTAACAGGAATCTCCAAACCTTAATAAAA 1505
|||||..... |||.....: |||..... |||.....
511 uSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGlyValVal 528
|||||..... |||.....: |||..... |||.....
1506 ATCTGCTAGTACCTGGAATAATCCATACCTCTTATCTGAGTAGTTC 1555
|||||..... |||.....: |||..... |||.....
528 eGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlyGlyLeu 544
|||||..... |||.....: |||..... |||.....
1556 TACAGATTAATAAAGGTTAGATTATTTATTTCTAAATAAAGAAAGATTA 1605
|||||..... |||.....: |||..... |||.....
545 CysAlaAlaLeuLysGluGlyCysPheTyrValAspHisSerGlyAla 561
|||||..... |||.....: |||..... |||.....
1606 TGTGTACCTTAAGAGAAATGCTGTTTATAGTACATATTCAGGGGC 1655
|||||..... |||.....: |||..... |||.....
561 aLeuArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArg 578
|||||..... |||.....: |||..... |||.....
1656 CATGAGAGACTCCCATGAACAAGCTTAATAAAGGTGAGAGAAACGTGAA 1705
|||||..... |||.....: |||..... |||.....
578 rGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSe 594
|||||..... |||.....: |||..... |||.....
1706 GGGAAAGAGAACTTACTCAAGATGTTTAAAGATGGTTCACAGAGTC 1755
|||||..... |||.....: |||..... |||.....
594 rProThrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValVal 611
|||||..... |||.....: |||..... |||.....
1756 TCTCTGTTGCTACCTACTTCTTACTTTAACGAGACCTTAATATCTC 1805
|||||..... |||.....: |||..... |||.....
611 eLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 627
|||||..... |||.....: |||..... |||.....
1806 TCTCTGCTGTTACTCAGAGTGGCCATGATATTATTAACAAGTAAATGCC 1855
|||||..... |||.....: |||..... |||.....
628 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnG1 644
|||||..... |||.....: |||..... |||.....
1856 TTCAATTAGAAACGAATATAGTCCAGATCATGTGTTACTTAGACACAA 1905
```



```

526 AATAAGCTTACCTTACTAGCTTGTGAGAAAGACACCTGATATAAAAA 575
374 salapropioSerHisGlnHisLeucystySerThrValValTyrGluG 391
576 GGTTCCTCCCATCCACCAACACCTTTACACCAACCACTGAACCTTTAATC 625
391 lna1aserGlnsngInTyrLeuValProGlyTyrAsnArgTPrtpala 407
626 AAACCTCTGAGAGTCAATATCTGGTACCTGTTATACAGAGTGGGGCA 675
408 CysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnse 424
676 TGTAAATACGTGATTACCCCTGTGTGTTCCACCTGGTTTCAACCAAC 725
424 rTysAspPheCysValMetValGlnIleValProArgValTyrTyrHisP 441
726 TAAAGACTTTTACATTATGTGTCCAAAATTTCTCCCGAGATATTTACTATC 775
441 roGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgProLys 457
776 CCAGAAACCAATCTCGATGATATGATTACAGAACCACTGCACAAAG 825
458 ArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThral 474
826 AAAAAACCATATCCCTGACACTGCAGTAATGCTCGAGACTCGGAGTAT 875
474 aValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGln 491
876 AACAGCTGTGAGAACAGCAAGCAAGCTTTAGTTACAGACTCGACGAC 925
491 euGlnLysGlyLeuGlyLeuHisAlaAlaMetThrGluAspLeuArg 507
926 TAGAAACAGGACTTAGTACCTACCTCAAAATTTGTAACAGAAACTCCCA 975
508 AlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSe 524
976 GCCCTAAAAAATCTGTGCTAGTACCTGAAAAATCCCTAACCTCCTATTC 1025
524 rGluValValLeuGlnAsnArgArgGlyLeuAspLeuPheLeuArg 541
1026 TGAAGTACTCTACAGAAATAAAAAGGTTAGATTATTATTCTAAAAA 1075
541 lueGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrValAsp 557
1076 AAAAGAGTTATGTGTACCTTAAAGGAGAAATGCTGTTTATGTAGAT 1125
558 HisSerGlyAlaAlaLeuArgAspSerMetAsnLysLeuArgLysLysL 574
1126 CATTCAGGGGCCCATCAGAGACTCATGACAGCTTAAAAAAGCTTGA 1175
574 uArgArgArgArgGluArgGluAlaAspGlnGlyTPrpPheGluGlyTPr 590
1176 GAAACGTGAGGAGGAAAAAGAACTTACTCAAAAGATGTTAAAAAGATG 1225
591 PheAsnArgSerProTPrpMetThrThrLeuLeuSerAlaLeuThrGlyPr 607
1226 TTCAACAGGTCTCCTGTGTGGCTAACCTTACTTCTTAAACAGGAC 1275
607 OleuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsn 624
1276 CTTAATAGTCTCCTCCTGTTACTACAGATGGGCGCATGATATATTACA 1325
624 rgrPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVal 640
1326 AGTAAATGTCCTTATTAGAAACGAATAGTGTGACGATCAGATCATGTGA 1375
641 LeuArgGlnGlnTyrGln 646
1376 CTTAGACACAGATACCA 1393

```

seq_name: /cgn2_6/ptodata/2/lna/6B.COMB.seq:US-09-011-745-7

```

seq_documentation_block:
; Sequence 7, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Joel
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 6312
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4058)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4059)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4060)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4061)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4247)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4248)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4249)
; OTHER INFORMATION: n is any nucleotide
US-09-011-745-7

alignment_scores:
Quality: 1508.50 Length: 746
Ratio: 3.060 Gaps: 24
Percent Similarity: 66.086 Percent Identity: 44.102

alignment_block:
US-09-171-553b-6 x US-09-011-745-7 ..
Align seg 1/1 to: US-09-011-745-7 from: 1 to: 6312

2 HisPro.....ThrLeuSerArgArgHisLeuProth 12
||||| ..||| ||||||| |||:
1428 CACCCGTTCCAGGTGCGGAGATCAAGTGTGTGTCAGACGCGCATGACCCAG 1477

```

```

12 r.....ArglyglGluProLysArgLeuarglle 23
   :      |||
1478 CAGGCTGAGCCTCGTGGAAGGCCATACCTGGTGTCTCATTACCC 1527
   |||
23 r.....AlaSerIleAlaTrpPhe.leuThrleuth 36
   :      |||
1528 CGACCGGGGTAAAGATGATGATGCTGCTGGGTCCATGCTCTCAC 1577
   |||
36 r.....IleThrProGlnAlaSerSerI 44
   :      |||
1578 CTCAAACCTGCACCCTTCGACACAGATGATGCT...GGAGCTGGA 1624
   |||
44 ysaArgLeuIle..... 47
   |||
1625 AAGAGCTGATCATCCTCTTAAGCTGGCTATTGGCGGGCGGAGAGCT 1674
   |||
48 ...AspSerSerAsnProHisArgProLeuSerLeuThrTrpLeuIle1 63
   :      |||
1675 CTGCAAAATTAAGAACCCCGACACGCCATGACCTCCTGGCAGGACT 1724
   |||
63 eaSPProAspThrGlyValThrValAsnSerThrArgGlyValAlaPro 80
   :      |||
1725 GTCC...CAACTGGAGACGTTGTCTGGATACAAAGGCGAGTCCAGCC 1771
   |||
80 rgllyThrTrpTrpProGluLeu.....HisPheCysIleuArgLeu... 93
   :      |||
1772 CTGGACTTGTGTGCCACACTTAACCTGATGTGTGCTTGGCGGCT 1821
   |||
94 .....IleAsnProAlaVally 99
   :      |||
1822 AGTCTGAGTCTGGGATATCCGGGAACCGATGTCGTCCTTAAGC 1871
   |||
99 sSerThrProProAsn..... 104
   :      |||
1872 AGTCAGACTCGGACTCAGACTACTGCGCTTAAGCAAAATCAGCT 1921
   |||
105 .....LeuValArgSer 108
   :      |||
1922 GGGAGCGCATAGGCTGACCTACCTCGGCTAGGACTAGAAATGGCAAGC 1971
   |||
109 TyrGlyPheTyrCysCysPro.....GlyThrGluLysGlu 121
   :      |||
1972 TCTACTTTACGATGTCCCGGATGGCGGACCTTCAGAGCTAG 2021
   |||
121 sTyrCysGlyGlySerGlyIleuSerPheCysArgArgTrpSerCysVal 138
   :      |||
2022 AAGGTGGGGGCTAGAAATCCTATAGTAAAGATGGGATGTGAGA 2071
   |||
138 hrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgVal 154
   :      |||
2072 CCACGGGACCGGTTATTGG...CTATCTAAATCTCCAAAGACCTCA 2118
   |||
2118 LysPheSerPheValAsnSerGlyProGlyLysTyrLysMetLysIle 171
   :      |||
2119 ACGTAAATATGGGACCAAAATAGCAGATGAGCTCAAAATTTACAGAG 2168
   |||
171 uTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLysIle 188
   :      |||
2169 TCACCAACCGGCTGTGTATACCC.....CTTAAATAG 2203
   |||
188 erPheThrGluArgLysThrGlyLysTyrSerLysValAspLysTrpTyr 204
   :      |||
2204 ATTTCAGACAAA.....GGAAATTTATCCAG.....GACTGGATA 2241
   |||
205 GluLeuGlyAsnSerPheLeuLeu.....TyrGlyGlyGlyAla 218
   :      |||
2242 ...ACGGGAAAAACCTGGGATTAAAGATTCTATGTGCTGGACATCCAG 2288
   |||
218 ySerThrLeuThrIleArgLeuArgIleGluThrGlyThrGluProPro 235
   :      |||
2289 CGTAGACTTCACCATTCGCTTAAAAATTCAC.....AACATCCAGCTG 2332

```

```

235 aAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 251
   |||
2333 TGGCAGTAGTCTCTGACCTCGTCCTGTGGAAACAAGACCTCTGAAGC 2382
   |||
252 Glu.....ProProHisAsnLeuProValProG 261
   :      |||
2383 TCCCTCGCTCTCCACCTCTCTCTCCCAAGGAGGCGCCACCCATC 2432
   |||
261 nLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSerThr 278
   :      |||
2433 TCTC.....CCGACTCTTACTCTCACAG 2455
   |||
278 hrGlyLeuIleProThrAsnThrProArg..... 287
   :      |||
2456 CCGTGGCGACTAGTGACAAACTCCACCGGTGAGAAAAAATTTGACC 2505
   |||
288 ...AsnSerProGlyValProValLysThrGlyGlnAlaGluPheSerIle 303
   :      |||
2506 CTAAACACTCCG.....CCTCCACACAGCGCAGACACTTTTGACT 2549
   |||
303 uIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThr 320
   :      |||
2550 TGTGAGGGGGCTTCTTAACCTTAATGCTACCAACCGAGGGCCACTG 2599
   |||
320 erSerCysTrpLeuCysLeuSerSerGlyProProTyrTyrGluGlyMet 336
   :      |||
2600 AGCTTGTGCTGCTTTGTTGGCCATGGGGCCCCCTTATTATGAACATA 2649
   |||
337 AlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGlnCysTh 353
   :      |||
2650 GCCTCATCAGAGAGGTGCTTACTCCACCGAC...CTTGACCGGTGCG 2696
   |||
353 rThrGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyT 370
   :      |||
2697 CTGGGGACCAAGAAAGCTACCTCAGAGGTCTTAGAGACAGGGGT 2746
   |||
370 hrCysIleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThr 386
   :      |||
2747 TGTGCATAGAAAGGTGCCCTTACCACATCAGACATCTCGCATAGAAC 2796
   |||
387 ValValTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAs 403
   :      |||
2797 CTATCCATCAATTCCTCCGAGACCATCATGATGCTCCCTCCACACA 2846
   |||
403 nArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrServ 420
   :      |||
2847 TAGCTGTGGGCTTGCAGACCTGGCTCACCCCTTGCTCTCCACTGAG 2896
   |||
420 aPheAsnGlnSerLysAspPheCysValMetValGlnIleValProArg 436
   :      |||
2897 TTTTATATCAGACTAGAGATTTCTGTATCCAGATCCAGTGTCTCTGC 2946
   |||
437 ValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyrArgTy 453
   :      |||
2947 ATCTATTAATCTATCTGAAGAAAGTTTGTACAGCCCTATGACAAATTC 2996
   |||
453 rAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuG 470
   :      |||
2997 CCCAGGACTAAAGAGAGGCTGTCTCACCTTACCTTACCTGTTTTACGTG 3046
   |||
470 LysLeuGlyThrAlaValAlaGlyValGlyThrGlyThrAlaAlaLeuThr 486
   :      |||
3047 GGTGGGAATCAGCGCGGGAATAGTACTGTTCACTGCTCTATTATTA 3096
   |||
487 GlyProGlnGlnLeuGlnLysGlyLeuGlyLysLeuHisAlaAlaMetTh 503
   :      |||
3097 GGACCTATAGACCTCCAGCAAGGCTGACAAACCTCCAGATCGCATAGA 3146
   |||
503 rGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGlnGluSerL 520
   :      |||
3147 TGTCACTCTCGGGCTTCAGAGCTCAGTCAGAAATTTAAGAGGCTCAC 3196
   |||
520 euThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 536

```

```

3197 TGACCTCCCTGTCGAGGTAGTGGCTCCAAAATAGAGAGGCTTGACTTG 3248
537 IeuphLeuArgIuArgIuYleuCyAlaIalaIeuLysGluIuGly 553
3247 CTGTTCCTAAAGAGAGGTGGCTCTGTGGCGGCGCTAAAGGAGAGTGGT 3298
553 sPheYrValAspHisSerIylAlaIleArgAspSerMetAsnIysLeuA 570
3297 TTTTTCATGTAGCCACACTCAGTGCAGTAGCGGAGCTCATGTAAGAAACTCA 3348
570 rGlyYsYsLeuGIuArgArGIuArgIuArgIuArgIuAlaAspGluGIuYr 586
3347 AAAAAAACGTGGATAAAAAGCAGTGAAGCGCGCGAAGAAAGCCAAACTGG 3396
587 PheGIuGIuYrPheAsnArgSerProTyrMetThrIhrLeuSerAl 603
3397 TTGGAAGGATGGTTCATATACTCCCTCGTTCCTACCTCGATATAC 3446
603 AleuThrGIuProLeuValIleuLeuLeuLeuLeuThrValGlyProC 620
3447 CATCGCTGGGCCCTATACCTCTCTCTGTTGCTATCTCGGGCAT 3496
620 ySLeuIleAsnArgPheValAlaPheValArgIuArgIuSerAlaVal 636
3497 GCATCATCAATCATTTAGTTCATTGGTTTAAACACAGATCTCAGTAGTC 3546
637 GluIleMetValLeuArgIuArgIuArgIuYrGluGIuYleu 648
3547 CAGGCTTAGTCTGCTGACTCACACATACACACAGCTA 3592
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1

```

```

seq.documentation.block:
: Sequence 1, Application US/08/16351A
: Patent No. 6033905
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based
: TITLE OF INVENTION: Retroviral Vectors
: NUMBER OF SEQUENCES: 5
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patenclin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/716.351A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/03784
: FILING DATE: 06-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 15280-128-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8535 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: 1..8535
: OTHER INFORMATION: /standard_name="Galv SPATO Genome"
: US-08-716-351A-1

```

[illegible]

Align seg 1/1 to: US-08-716-351A-1 from: 1 to: 8535

[illegible]

```

6545 TCACGACGCGGCTGGTGAACCC.....CTTAATAAG 6579
188 ePheThrGluArgLysThrGlyLysTyrSerLysValAspLysTyr 204
6580 ATTTCACAGACAAA.....GGAAAAATTATCCAG.....GACTGGATA 6617
205 GluLeuGlyAsnSerPheLeu.....TyrGlyGlyGlyAla 218
6618 ...ACGGGAAAAACCTGGGGATTAGATTCTATGTGTGGACATCCAG 6664
218 ySerThrLeuThrIleArgLeuArgIleGluThrGlyThrGluProPyr 235
6665 CGTACAGTTCCACCATTCGCTTAATAATCAC.....AACATCGACGCTG 6708
235 aAlaMetGlyProAspLysValLeuAlaGluGluGlyProAlaLeu 251
6709 TGGCAGTAGGCTCTGACCTCGCTCTGTGGAGAACAGACCTCTGGAACG 6758
252 Glu.....ProPheAsnLeuProValPro 261
6759 TCCCTGCTCTCCACCTCTCTCTCCCAAGGAGGCGCCACGCCCATC 6808
261 nLeuThrSerLeuArgProAspIleThrGluProPheSerAsnSerThr 278
6809 TCTC.....CCGACTCTACTCCACAG 6831
278 hGlyLeuIleProThrAsnThrProArg..... 287
6832 CCGTGGCGACTAGTCACAACTCCACGGTGAGAAAAACAATTGTAC 6881
288 ...AsnSerProGlyValProValLysThrGlyGlnArgLeuPheSer 303
6882 CTAAACACTCG.....CTCCACACACAGGCGAGACATTTTGATCT 6925
303 uIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThr 320
6926 TGTGAGGGGGCTTCTTAACCTTAAGCTAACACAGGCGGCACTG 6975
320 eSerCysThrLeuCysLeuSerSerGlyProPyrTyrGlyGluMet 336
6976 AGCTTGTGGCTTGTGTGGCCATGGGCCCCCTTAATATGAAACAATA 7025
337 AlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGlnCysTh 353
7026 GCGTATCAGAGAGAGTGGCCCTACTCCACCGAC...CTTGACGGTGGCG 7072
353 rTrpLysSerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyT 370
7073 CTGGGGGAGCCCAAGAAAGCTCACCCTCAGAGGTCTCAGAGACGGGT 7122
370 hCysIleGlyLysAlaProPheSerHisGlnHisLeuCysTyrSerThr 386
7123 TGTGATGAGGAAGGTGCCCTTACCCTAGCATCTGTGCAATCAGACC 7172
387 ValValTyrGluGlnAlaSerGlnAsnGlnTyrLeuValProGlyTyrAs 403
7173 CTAATCATCAATTCCTCCGGAACCATAGTATGCTGCCCTCCACCA 7222
403 nArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSer 420
7223 TAGCTGTGGGCTTGCACAGCTCAGCCCTGCTGCCCTCCACCTCAG 7272
420 aPheAsnGlnSerLysAspPheCysValMetValGlnIleValProArg 436
7273 TTTTAACTAGACTAGATTTCTATATCCAGGTCCACCTATTCCTGCTGC 7322
437 ValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyrArgTy 453
7323 ATCTATTAATCTATCTGAGAAAGTTTGTGTACAGGCTATGACAATTCCA 7372
453 rAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuG 470

```

```

7373 CCCGAGGACTAAAGAGAGGCTGTCTCATTACCTAGCTGTTTACTGG 7422
470 lYleuGlyThrAlaValGlyValGlyThrGlyThrAlaIleThr 486
7423 GGTGGGGAATACGCGGGGAATAGTACTGTTCACTGCTTAATTA 7472
487 GlyProGlnGlnLeuGlyLysGlyLeuGlyGluLeuHisAlaAlaMetTh 503
7473 GGACCTATAGACTTCACGACAGGCTGACAGAGCTCCAGATGCGCATAGA 7522
503 rGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluSerL 520
7523 TGTCACTCCGCGGCTCCCAAGACTCAGTCAGCAAGTTAGAGGACTCAC 7572
520 eThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 536
7573 TGACTTCCCTCTCCGAGTATGCTCCAAATAGAGAGGCTTGACTTG 7622
537 LeuPheLeuArgGluGlyLeuCysAlaAlaLeuLysGluGlyCys 553
7623 CTGTTCTAAAGAAGGTGGGCTCTGTGCGGCTTAAGAGAGAGTCTG 7672
553 sPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeu 570
7673 TTTTACATAGACCACTCAGTGCAGTACGAGACTCATGAAAAACTCA 7722
570 rGlyLysLysLeuGluArgArgArgGluArgGluAlaAspGlnGlyTrp 586
7723 AAGAAAACTGGATTAAGACAGTTAGAGCGCCGAAACCAAACTGG 7772
587 PheGlnGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAl 603
7773 TATCAGAGATGTTCAATTAACCTCCCTGTGTGCTACTACCTGCATAC 7822
603 aLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGlyProC 620
7823 CATCGTGGGCGCCCTATTAACCTCTCTGTGTGCTCATCTCGGGCCAT 7872
620 ySerLeuIleAsnArgPheValAlaPheValArgGluValSerAlaVal 636
7873 GCATCATCAATTAAGTTAGTCAATTCATCAATATAGATAGTCAATGT 7922
636 lGlnIleMetValLeuArgGln 643
7923 TAAATTTCTGCTCTTGAACAA 7944

seq_name: /cgn2_6/plodata/2/lna/6A.COMB.seq:US-09-075-272-1
seq_documentation_block:
; Sequence 1, Application US/09075272
; Patent No. 6136598
; GENERAL INFORMATION:
; APPLICANT: MILLER, A. DUSTY
; APPLICANT: WOLGAMOT, GREG
; APPLICANT: BONHAM, LYNN
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
; TITLE OF INVENTION: PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/075,272
; FILING DATE: 08-MAY-1998

```


414	rocyvalSerThrSerValPheasnGlserysAspPheCysValmet	430
:	: : : : :	
7060	CTTGATATCTATGTCGTGTTTCAATAAGCTCCAAAGATTTCTGTATTTTA	71099
431	ValGlnIleValProArgValTYrTYrHisProGluIuValValLeuAs	447
7110	GTTTCAGCTTATTCCTAGTACCCTCGTATCATGATGATGATGCATTTCTAGA	71599
447	pGluTYrAspTYrArgTYrAsnArgProIlysArgIuProValSerLeuT	464
:	: : : : :	
7160	CAAAATTTGAACATCGGGTC...CGCTGGAAAGAAAGAACCCATTACTTTAA	72060
464	hIleuAlaIaIaMetLeuGlyLeuGly...ThrAlaValGlyValGlyThr	479
7207	CGTTAGCACTGCTTTTAGATGTTGGAGTGCGCCGACGTGGGGTAGTACT	72560
480	GlyThrAlaIaIaLeuIleThrGlyProGlnGluLeuGluIuysGlyLeuG	496
:	: : : : :	
7257	GGAACCGCTGCTTAATCCAGACCCCCCGATACTTTAG.....	72950
496	yGluLeuHisAlaIaIaMetThrGluIuAspLeuArgAlaLeuLysGluSerV	513
7296	.GAATTACTAGTACGTATGATGATCTGATCTCAGAGCATATGAAACACTTA	73444
513	aISerAsnLeuGluIuGluSerLeuThrSerLeuSerGluValValLeuGln	529
7345	TAAACAAACTAGAGAGATCTTTAACTCTCTGCTGTGGAGTAGTACTCAA	73944
530	AsnArgArgGlyLeuAspLeuLeuPheLeuArgGluIuGlyLeuCysAl	546
7395	AATAGGAGAGGATTAATTTGTTATTTCTTAAGACAGAGAGACTTTGTGC	74444
546	aAlaLeuLysGluIuLysCysPheTYrValAspHisSerGlyValaIleA	563
7445	TGCTCTCAAGAGAGATGTTGTTTATATGTCGACACTCGGAGATGATCA	74944
563	rAspSerMetAsnLysLeuArgLysLysLeuGluATGATGATGATGArgGlu	579
7495	AAGATTCTATGGCCAAACTTAGAGAACCGCTAGTATATACGTCAAAGAGC	75444
580	ArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProT	596
7545	AAGAGAGCAACAAGAGATGTTTGAAAGCTGTTTATATAGTCCCTCG	75944
596	pMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuL	613
7595	GCTCACCACTCTCTCTCCACTATAGGGGAGCCCTTATATTAATCTTGCC	76444
613	euleuleuThrValGlyProCysLeuLLeasnArgPheValAlaPheVal	629
7645	TTTTGCTTACTTTGGCCCTCGATCTTAAATAGTAGAGCCCTTATAT	76944
630	ArgGluArgValSerIaValGlnIleMetValLeuArgGlnGlnIuTYrG	646
7695	AGAGAAAGATTAATATCATCAGTACAGGTTATGTGATTTAAACACAAATATCA	77444
646	nGlyLeuLeuSerGlnGlyGluThrAspLeu	656
7745	G...GTCTTCAGAGAGGCCGCAAAATCTGCCTC	7772
seq_name: /cgn2_6/ptodate/2./ina/5A.COMB.seq:US-08-110-300A-8		

STREET: 45 Rockefeller Pl.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/110,300A
FILING DATE: 20-AUG-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Jacobs, Seth H.
REGISTRATION NUMBER: 32,140
REFERENCE/DOCKET NUMBER: 11698A50
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-757-2200
TELEFAX: 212-586-1461

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 8323 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

US-08-110-300A-8

```

alignment_scores:      Length: 709
                        Quality: 1299.00
                        Ratio: 2.855
                        Gaps: 26
Percent Similarity:    64.175
                        Percent Identity: 42.595

alignment_block:
US-09-171-553B-6 x US-08-110-300A-8 ..

```

```
seq_name /cgn2_6/p1codata/2/ina/5A.COMB.seq:US-08-11
:
: seq documentation block :
: Sequence 8, Application US/08110300A
: Patent No. 5643756
:
: GENERAL INFORMATION:
:
: APPLICANT: Pfizer, Abraham
: APPLICANT: Kayman, Samuel
: TITLE OF INVENTION: FUSION GLYCOPROTEINS
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Davis Hoxie Faithfull and Hapgood
```

```

1      ign seg 1/ to: us-08-110-300A-8 from: 1 to: 8323
2
3      4   ThleuSerArgArgHisIleuProThrArgGlyGlyIuProArgyle 20
4      5787 ACcCTCTCAAAATCC.....CCTAAAGATTAAGATTGACCGCGGACCT 5830CGG
5      20 uArgIleProLeuSerPheAlaSerIleAlaIrrPheLeuThrIleuThrI 37
6      5831 CCTAAATCCCTTAATTC.....TTCCGTGTCTC... 5861GGT
7      37 IethrProGAlaSerIlyArgLeuIleAspSerSeruProHis 53
8      5862 .....AAAGGGCCAGATCCGCGACGACCCGGCTCCAGCCCTCAC 5900CTC
9      54 ArgProLeuSerIleuThrIleuIleIleAspProAspThrGlyValTh 70
10     5901 CAGGTGTACAACTTACTCTGCGGAGAGACCAATGGCGATCGG...GAGAC 5947G
11     70 rValIasSerThrArgGlyValAlaIrrArgGlyIlyThrPrrProIuL 87
12     5948 AGATGGGCAATTCAGGCAACACCCTCTGTGCACTGTGGTGGCCAGTCC 5992G
13     87 eu.....HisPhe... 89
14     5998 TCACCCAGATTGTGTATGTATGCTCTCAGTGGGCCGCCGCCACTGGGGG 6047G
15     89 ..... 89
16     6048 CTAGAGATCAGGCCCTTATTCCTGACCCCGGGGCCCTTGTGTGTC 6097G
17     90 .....CysIleuArgIleuIleAsnProAlay 98
18     90 ..... 98

```

```

6098 AGGAGACAGCGGAGACGTTGCAGCTGTGCGACAGACTGCAACGAGCCCT 6147
98  allyserThrProPro.....AsnLeuValArg... 107
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6148 TGACCTCCTCCTACCCCTCGGTGCACACACTGCTGAGACACTTAAGCTG 6197
108 .....SeryTyrGlyPheTyrCysCysProGlyThr 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6198 GACCAGGTAACTCATTAATCACTGAGGATTTTATGCTGCCCGCGGTC 6247
117 rgluLys.....GluLysTyrCysGlyGlySerGlyGlySerPheC 131
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6248 ACATCGCCCGCCGGGAGAGCAAGCTGTGGGGTCCAGACTCCTCTACT 6297
131 ysArgrArgrPserCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6298 GTGCCCTCTGGGGGTGCGAGACACCGGTAGAGTATACGGAAG...CCC 6344
148 lIeSerLeuGlnAspArgValLysPheSerPheValAsnSerGlyProG1 164
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
6345 TCCTCTTCTGGACTACATCACA.....GTAGACACAACTCTAC 6385
164 yLysTyrLysMetLysLeuTyrLysAsp...LysSerCysSerProS 180
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6386 CTCTAACACAGCTGTTCAGGTATGCAAGACAAATAGTGTCATCC. 6434
180 eAspLeuAspTyrLeuLysIleSerPheThr.GluArgLysThrGlyL 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6435 .....TTGGCTATCCGGTTTACAAACGCCGCGGAAACAGGTC 6470
196 ySTySerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeu.Le 212
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6471 ACCTC.....ATGCACACTGACACATATGCGGTCTACGCTTATGT 6514
212 uTyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluT 229
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6515 CTCTGACAGAGACCAGGCTTACTTTCGGATCCGACTCAGTTTCA. 6563
229 hGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 245
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6564 ..AACTAGGACCTCGATCCCAATAGGACCAACCCGCTCGACAGAC 6611
246 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGlnLe 262
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
6612 CA.....CTTTCGTTCCCGCTACCTAATCCCTACCC..... 6644
262 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrG 279
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6645 .....AAACTGCCAAGTCTCCCGCCCTCTAGTTGACTGCCA 6684
279 lYleuIle...ProThrAsnThrProArgAsnSerProGlyValProVal 294
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
6685 CATTAATTTCCCGTCCCGCCACTCCACTCCAGCCCG.....CCAGCA 6728
295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaI1 311
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
6729 GGAAGGGAGACAGATTACTAATTAATTAAGGAGCTTACAGGCACT 6778
311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSers 328
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6779 CAACCTTACCAACCTGATTAACACTCAGAGTGGCTTATGCTAGCT 6828
328 eArgLysProProTyrTyrGluGlyMetAlaLysGluAlaGlyLysPheAsnVal 344
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
6829 CTGACCCCGCCCTATTACGAGGGGTGGCTCTAGACTTATTCACAC 6878
345 ThrLysGlnHisArgAsnGlnCysThrTrpLysSerArgAsnLysLeuTh 361
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6879 CATACTGTGCCCAAGCTAACTGCTCGTGCTCCCAACACAGGCTGAC 6928
361 rLeuThrGlnValSerGlyLysGlyThrCysIleGlyLysAlaProProS 378
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
6929 CCGTCGCAAGTGAAGTGAAGGAGCTCTGATGAGAACAGTCCCAAAA 6978

```

```

378 eThsGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGlu 394
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6979 CTCACAGCCCGCTGTGCAACACTACCTT.....AAGGACGCAAGG 7022
395 AsnGlnIleuValProGlyTyrAsnArgTrpTrpLysAsnThrG1 411
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
7023 TCTTACTATCTAGTGTGCCCCACAGGACATATGTGGCCATCTTAACCTGG 7072
411 lYleuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheC 428
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7073 ACTACTCTCAAGCCCTATGTGCAACCGTCTTAATTCGACACACTACTATT 7122
428 yValMetValGlnIleValProArgValTyrTrpHisProGluGlnVal 444
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7123 GCGTTCGTGTGAATATATGCGCCAGGGTCACTACATCCCTCCAGTTAC 7172
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgLysProVa 461
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7173 GTCTATACCGAGTTTGAA...AAATCCATAGACATTAAGAAGAACAGT 7219
461 lSerLeuThrLeuAlaValMetLeuGly.....LeuGlyThr 474
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7220 GTCCCTTAACCTTGCCCTTATTTAGTGGGCTACTATGGTGCAATCG 7269
474 lValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGln 490
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7270 CCGCGGAGATGAGGAGAGAACTACCGCCCTGGTC...GCCACCAACAG 7316
491 LeuLysLysGlyLeuGlyGluLeuHisAlaAlaMetThrGlnLysLeuAr 507
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
7317 TTTCAG.....CAGCTCCATGCTGCGGTACACAAATGATCTCAA 7354
507 gAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuS 524
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
7355 AGAAGTGGAAAGTCAATTAATTAACCTAGAAAGTCTTACTTCTGTTG 7404
524 eArgLysValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArg 540
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7405 CTGAGGTTGTAATGACAGATCGACAGAGGCTAGACGTTGTTCTTAATA 7454
541 GlnGlyGlyLeuCysAlaAlaLeuLysGluGlyCysCysPheTyrValS 557
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7455 GAGGAGAGACTGTGTGCTGCTTAAGAAAGATGTTTCTATCTGA 7504
557 PHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLeuG 574
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7505 CCATACAGGCTAGTAGATAGTATGGCCAAATTAAGAGAGACTCT 7554
574 lUArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrp 590
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
7555 CTCAGAGACAAAACTATTGAGTCAAGCCCAAGATGTTGCAAGATGG 7604
591 PheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrLysTr 607
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7605 TTTTAAAGATCCCGCTGTGTTTACACGTGATATTCACCATCATGCGGGCC 7654
607 oLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnA 624
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7655 TCTCATTTATCTCTACTAATTTCTGTTTGGACCTGATTTCTTAATC 7704
624 rGPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVal 640
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7705 GATTAGTCAATTGTTTAAAGACAGATCTCAGTACTCCAGGCTTATGTC 7754
641 LeuArgGlnGlnTyrGlnGlyLeu 648
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
7755 CTGACTCAACATATACCAACAGCTA 7778

```

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq.us-08-886-642-8
seq_documentation_block:
; Sequence 8, Application US/08886642

Patent No. 5952474
 GENERAL INFORMATION:
 APPLICANT: Pinter, Abraham
 APPLICANT: Kayman, Samuel
 TITLE OF INVENTION: FUSION GLYCOPROTEINS
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 45 Rockefeller Plaza, Suite 2800
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/886,642
 FILING DATE: 01-JUL-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/110,300
 FILING DATE: 20-AUG-1993
 APPLICATION NUMBER: 07/938,100
 FILING DATE: 28-AUG-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Hone, William J.
 REGISTRATION NUMBER: 26,739
 REFERENCE/DOCKET NUMBER: 0763/010002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212/765-5070
 TELEFAX: 212/258-2291
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8323 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 US-08-886-642-8

alignment_scores:
 Quality: 1299.00 Length: 709
 Ratio: 2.855 Gaps: 26
 Percent Similarity: 64.175 Percent Identity: 42.595

alignment_block:
 US-09-171-553b-6 x US-08-886-642-8 ..
 Align seg 1/1 to: US-08-886-642-8 from: 1 to: 8323

4 ThrLeuSerArgArgHisLeuProThrArgGlyGluProLysArgLe 20
 |||||
 5787 ACGCTCTCAAAATCC.....CTTAAGATAGATTGACCCGGGAGACT 5830
 20 uArgTLeuProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeu 37
 |||||
 5831 CCTAATCCCTTAATCTC.....TTCCGTCTCTC.... 5861
 37 LeuThrProGlnAlaSerLysArgLeuIleAspSerSerAsnProHis 53
 |||||
 5862AAAGGGCCAGATCCGCGACGCCGCGCTCCAGCCTCAC 5900
 54 ArgProLeuSerLeuThrTrpLeuIleIleAspProAspTrpGlyVal 70
 |||||
 5901 CAGCTCTAACAATCTACGTGGGAGTGCACCAATGGGATCG...GAGAC 5947
 70 ValAlaSerThrArgGlyValAlaProArgGlyThrTrpTrpProGln 87
 |||||
 5948 AGTATGGCAATATTCAGGCAACCAACCCCTGTGGACTGTGGCCAGTCC 5997

87 eu.....Hispe... 89
 |||||
 5998 TCACCCAGATTGTGTATGTAGCTCTGAGTGGGCCGCCCACTGGGG 6047
 89 89
 6048 CTAGAGTATCAGGCCCCCTATCTCGCCCCGGGGCCCTTGTGCTC 6097
 90CysLeuArgLeuIleAsnProAlaVal 98
 6098 AGGAGACGGCGGAGACCTTGACAGCTGTGCCAGAGACTGCAGACCC 6147
 98 aLysSerThrProPro.....AsnLeuValArg... 107
 |||||
 6148 TGACCTTCCCTACCCCTCGGTGCACACTGCTCGAAGACTTAAGCTG 6197
 108SerTyrGlyPheTyrCysCysProGln 117
 |||||
 6198 GACCGAGTACTATATAATCAAGTGGATTTATGTCTGCCCGCGCTC 6247
 117 rGluLys.....GluLysTyrCysGlyGlySerGlyLysSerPheC 131
 |||||
 6248 ACATCGCCCGCGGAGACCAAGTCTGTGGGGTCCAGACTCTTCTACT 6297
 131 ysArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrp 147
 |||||
 6298 GTGCCCTTGGGGCTGGCAGACCAACCGGTAGAGTATCTGAGAG...CC 6344
 148 IleSerLeuGlnAspArgValLysPheSerPheValAlaSerGlyPro 164
 |||||
 6345 TCCTCTTCTTGGGACTCATACACA.....GTAGCAACATCTCAC 6385
 164 yLysTyrLysMetLeuLysLeuTyrLysAsp...LysSerCysSerPro 180
 |||||
 6386 CTCTAACCAAGCTTTCAGGTATCCAAAGCAATAGTGTGCAATGCC. 6434
 180 eAspLeuAspTyrLeuLysIleSerPheThr...GluArgLysThrGly 196
 |||||
 6435TTGGCTATCCGTTTACAAACCGCGGGAACACAGCTC 6470
 196 yTyrSerLysValAlaLysTrpTyrGluLeuGlyAsnSerPheLeu 212
 |||||
 6471 ACCTC.....ATGGACAACCTGGACACTATTGGGGCTTCTTATGT 6514
 212 uTyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGlu 229
 |||||
 6515 CTCTGGACAGGACCCAGGCTTACTTTCGGATCCGATCAGTATGAA. 6563
 229 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAla 245
 |||||
 6564 ...AATCTAGACACTCGATCCCATAGACCAACCCGCTCTCGCAGAC 6611
 246 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGln 262
 |||||
 6612 CAA.....CTTCGTTCCCGTACTATATCCCTACCC.... 6644
 262 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrG 279
 |||||
 6645AAACCTGCCAAGTCTCCCGCGCTAGTTGCAGCTCCA 6684
 279 LysLeuIle...ProThrAsnThrProArgAsnSerProGlyValPro 294
 |||||
 6685 CATGTATTTCCCGTCCCGCTCCACTCCACTGAGCCCG...CCAGCA 6728
 295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAla 311
 |||||
 6729 GGAACGGGAGACAGATTACTTAATCTAGTACAGGAGCTTACCGGACT 6778
 311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuGlyLeuSer 328
 |||||
 6779 CAACCTTACCAACCTGATAAAGTCAAGAGTGTGCTTAACTGCTAGTGT 6828
 328 eArgLysProProTyrTrpGlyGlyMetAlaLysGluArgLysPheAsn 344

```

6829 CTGAGACCCCTATTAGAGGGGTTGCCCTCTAGTACTATTTCACAC 6878
345 ThLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeu 361
6879 CATACCTCTGCCACACTACTCTCCGTGGCTCCCAACACAGCTGAC 6928
361 rLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProP 378
6929 CCTGTCCGAGTGTACTGACGGGGCTGTCATAGGACAGGCCAAAAA 6978
378 erHisGlnHisLeuCysTyrSerThrValValTyrGlnGlnHisSer 394
6979 CTCACAGAGGCCCTGTGCACACACTACCTT...AAGCAGAGCAAGGG 7022
395 AsnGlnTyrLeuValProGlyTyrAsnArgTrpTyrAlaCysAsn 411
7023 TCTTACTATCTAGTTGCCCCACAGAACATATGCGCATATACCTGG 7072
411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAsp 428
7073 ACTCACTCATGCTATCTGCGACCGCTGTAATCGCACACTGACTATT 7122
428 yValMetValGlnIleValProArgValTyrTyrHisProGlnGlu 444
7123 GCGTCTCGTGAATATGCGCAGGTCCTACCTACCATCTCCCGATTAC 7172
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGlu 461
7173 GTCCTATAGCCAGTTGAA...AAATCCATATGACATAAAGAGACAG 7219
461 lSerLeuThrLeuAlaValMetLeuGly...LeuGlyThr 474
7220 GTCCTTAACCTTGCCCTTATTATTAGGGCTTAACATGAGGTGCGAT 7269
474 lValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGln 490
7270 CCGCGGAGTAGGAGACAGAACTACCGCGCTGTC...GCCACCCAGAG 7316
491 LeuGluLysGlyLeuGlyGlnHisAlaAlaMetThrGluAspLeu 507
7317 TTTTCAG...CAGCTCCATGCTGCGGACAGATGATCTCA 7354
507 gAlaLeuLysGlnSerValSerAsnLeuGlnGlnSerLeuThrSer 524
7355 AGAAGTCGAAAAAGTACTTACTACCTGAAAAGCTCTTCTTCGTGT 7404
524 erGluValValLeuGlnAsnArgArgGlyLeuAspLeuPheLeuArg 540
7405 CTGAGGTGTACTGCGAATCGACGAGGCTAGACCTGTCTCTTCTAAA 7454
541 GluGlyGlyLeuGlyAlaAlaLeuLysGlnLysCysPheTyrVal 557
7455 GAGGAGAGACTGTGTGCTGCCCTAAAGAAAGATGTTCTTCTGTCTGA 7504
557 pheSerGlyAlaIleArgAspSerMetLysLysLeuArgLysLys 574
7505 CCATACAGCGCTAGTAAAGATTAATGCCCAATTAAGAGAGAGACTCT 7554
574 luaArgArgArgArgGluArgGluAlaAspGlnGlyTyrPheGlnGly 590
7555 CTCAGAGACAAACATTTTGAGTCGAGCCAAAGATGTTCTGAAGATG 7604
591 PheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGly 607
7605 TTTTACAGATCCCTGCTGTTTACACGCTGATATCCACATCATGAGG 7654
607 OlauValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIle 624
7655 TCTCATATATCTCTACTAATTCGCTTTTGGACCTGCATCTTATATC 7704
624 rGpIleValAlaPheValArgGluArgValSerAlaValGlnIleMet 640

```

```

7705 GATTAGTCAATTGTTTAAGACAGATCTCAGTAGTCAGGCTTACTC 7754
641 LeuArgGlnGlnTyrGlnGlyLeu 648
7755 CTGACTCAACATATACACACAGCTA 7778

seq_name: /cgn2.6/prodata/2/lna/pctus_comb.seq:pct-US93-08041-8
seq_documentation_block:
  ; Sequence 8, Application pct/US9308041
  ; GENERAL INFORMATION:
  ; APPLICANT: The Public Health Research Institute of the City of
  ; APPLICANT: New York, Inc.
  ; TITLE OF INVENTION: FUSION GLYCOPROTEINS
  ; NUMBER OF SEQUENCES: 16
  ; CORRESPONDENCE ADDRESS:
  ; ADDRESSEE: Davis Hoxie Faithful and Hapgood
  ; STREET: 45 Rockefeller Pl.
  ; CITY: New York
  ; STATE: New York
  ; COUNTRY: USA
  ; ZIP: 10111
  ; COMPUTER READABLE FORM:
  ; MEDIUM TYPE: Floppy disk
  ; COMPUTER: IBM PC compatible
  ; OPERATING SYSTEM: PC-DOS/MS-DOS
  ; SOFTWARE: PatentIn Release #1.0, Version #1.25
  ; CURRENT APPLICATION DATA:
  ; APPLICATION NUMBER: pct/US93/08041
  ; FILING DATE:
  ; CLASSIFICATION:
  ; PRIOR APPLICATION DATA:
  ; APPLICATION NUMBER: US 07/938,100
  ; FILING DATE: 28-AUG-1992
  ; PRIOR APPLICATION DATA:
  ; APPLICATION NUMBER: US 08/
  ; FILING DATE: 20-AUG-1993
  ; ATTORNEY/AGENT INFORMATION:
  ; NAME: Jacobs, Seth H.
  ; REGISTRATION NUMBER: 32,140
  ; REFERENCE/DOCKET NUMBER: 11698A50
  ; TELECOMMUNICATION INFORMATION:
  ; TELEPHONE: 212-757-2200
  ; TELEFAX: 212-586-1461
  ; INFORMATION FOR SEQ ID NO: 8:
  ; SEQUENCE CHARACTERISTICS:
  ; LENGTH: 8323 base pairs
  ; TYPE: nucleic acid
  ; STRANDEDNESS: double
  ; TOPOLOGY: linear
  ; MOLECULE TYPE: DNA (genomic)
  ; HYPOTHEICAL: NO
  ; ANTI-SENSE: NO
  ; PCT-US93-08041-8

alignment_scores:
  quality: 1299.00      length: 709
  ratio: 2.855          gaps: 26
  percent similarity: 64.175  percent identity: 42.535

alignment_block:
  US-09-171-553b-6 x pct-US93-08041-8 ..
  Align seg 1/1 to: pct-US93-08041-8 from: 1 to: 8323

4 ThLysSerArgArgHisLeuProThrArgGlyGluProLysArgLe 20
5787 ACAGCTCTCAAAATCC...CTTAAGATTAAGATGACCGCGGAGCT 5830
20 uArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrIle 37
5831 CCTATCCCTTAATTC...TTCCGTCTCTC... 5861

```

37 leThnProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHis 53
5862 AAAGGGCCAGATCCGACGACACCCGGCTCCAGCCCTCAC 5900
54 ArgProLeuSerLeuThrTrpLeuIleLeAspProAspThrGlyValTh 70
5901 CAGGCTACAACTTACTCTGGGAGTAGTACCAATGGGATCGG...GAGAC 5947
70 rValAsnSerThrArgGlyValAlaProArgGlyThrTrpTrpGlnLut 87
5948 AGATAGGCGCAATATCAGCAACCCCTGTGTGACTTGGTGCCAGTCC 5997
87 eu.....Hispe... 89
5998 TCACCCCAATTGTGTATGTAGCTCTAGTGGCGCCCACTGGGG 6047
89 89
6048 CTAGAGTACAGGCCCTATTCTCGGCCGGGCCCTGTGTGCTC 6097
90CysLeuArgLeuIleAsnProAlaVal 98
6098 AGGAGACAGCGGGAACGTTGCAGGCTGTGCCAGAGACTGCACAGACCT 6147
98 allYserThrProPro.....AsnLeuValArg... 107
6148 TGACCTCCTCACCCCTCGTGCAACACTGCTGGAACAGACTTAACCTG 6197
108SerTyrGlyPheTyrCysCysProGlyTh 117
6198 GACCAAGTACATCAATTAATCAAGTAGGAGATTATGTGCCCCGGCTC 6247
117 rGluLys.....GluLysTyrCysGlyGlySerGlyLysSerPheC 131
6248 ACATCCGCCCGGGAAGCCAGTCTCTGGGGGTCCAGACTCCTTACT 6297
131 yAsnArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
6298 GTGGCTCTTGGGGGTGGAGACACCGGTAGATATCTGGAG...CCC 6344
148 lIleSerLeuGlnAspArgValLysPheSerPheValAsnSerGlyProGln 164
6345 TCCTCTCTTGGGACTACATCA.....GTACACAACAATCCAC 6385
164 yLysTyrLysMetMetLysLeuTyrLysAsp...LysSerCysSerPro 180
6386 CTCATACCAAGCTGTTCAGATGCAAGACAAATAGTGTGCATGCC. 6434
180 eAspLeuAspTyrIleuLysIleSerPheThr...GluArgLysThrGly 196
6435TTGGCTATCCGGTTTCAAAAGCGCGGAACACAGTCC 6470
196 ystYrSerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeu 212
6471 ACCTC.....ATGGACAACCTGACACTATTGGGGTCTACGCTTTATGT 6514
212 uTyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluT 229
6515 CTCGTGACAGGACCGAGGCTTACTTTCGGATCCGACTCAGTTATCAA. 6563
229 hncLysThrLupProValAlaMetGlyProAspLysValLeuAlaGlu 245
6564AATGTAGACCTCGGATCCCAATAGACCAAAACCCGTCCTGGCAGAC 6611
246 GlnGlyProProAlaLeuGlnLupProHisAsnLeuProValProGlnLe 262
6612 CA.....CTTTCGTTCCCGCTACCTAATACCTCCTACCC..... 6644
262 uThrSerLeuArgProAspLleThrGlnProProSerAsnSerThrThrG 279
6645AAACCTGCCAAGTCTCCCGCGCTCTAGTGCAGCTCCA 6684
6685 CATTGATTTCCCGTCCCGCTCCACTCCACTCAACCCCGG.....CCAGCA 6728
295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAla 311
6729 GGAACGGAGAGACAGATTACTAATATAGTACAGGAGGCTTACAGGCACT 6778
311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysIleSer 328
6779 CAACCTTACCAACCCCGATTAACCAAGAGTGGTGTATATGCTAGTGT 6828
328 eArgLupProProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnVal 344
6829 CTGGACCCCGCTATTACAGAGGGGTTGCCGTCTAGTACTTATTCACAC 6878
345 ThrLysGlnHisArgAsnGlnCysThrTrpTrpGlySerArgAsnLysLeuTh 361
6879 CATACCTCTGCCACAGCTAAGTCTCCGGGCTCCCAACAAGACTGAC 6928
361 rLeuThrGlnValSerGlyGlyThrCysIleGlyLysAlaProPro 378
6929 CTGTCCGAGAGTACGTGACAGGGGACTGCTGATTAAGGAACAGTCCCAAAA 6978
378 eHisGlnHisLeuCysTyrSerThrValTyrGluGlnAlaSerGlu 394
6979 CTCACAGCGCCCTGTGCACACTACCTT.....AAGCAGGCAAGGG 7022
395 AsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGln 411
7023 TCTTACTATCTAGTCTGCCCCACAGAGAACTATGTTGGCATGTACACATCG 7072
411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheC 428
7073 ACTCATCCAGGCTTATCTGCCACAGTGGTATATGCAACACTGATATT 7122
428 yValMetValGlnIleValProArgValTyrTyrHisProGlnIleVal 444
7123 GCCTTCGTGGAATTATGAGCCAGGCTCACTACCATCTGCCAGTTAC 7172
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProVal 461
7173 GTCTATGACCAAGTTTGA.....AAATCCATGACATTAAGAAGAACCCAGT 7219
461 lSerLeuThrLeuAlaValMetLeuLys.....LeuGlyThr 474
7220 GTCTTATACCTTGCTTATATTAAGTGGCTAAGTATGGGTGGCATCG 7269
474 lValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGln 490
7270 CCGGGGAGTAGGACAGACACTACCGCTGTGC...GCCACCCAGAG 7316
491 LeuGlnLysGlyLeuGlyGluLeuHisAlaAlaMetThrGlnAspLeuArg 507
7317 TTTGAG.....CAGCTCCATGCTGGCTGCAAGATGATCTCAA 7354
507 gAlaLeuLysGluSerValSerAsnLeuGlnLupSerLeuThrSerLeu 524
7355 AGAAGTGAAGAAAGCTAATTAACCTTAAGAAAGTCTTACTTCTGTGT 7404
524 eArgLupValAlaLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArg 540
7405 CTGAGGTTGTAGTCGCAAAATCGACGAGGCTAGACTGTGTTCCTAATA 7454
541 GlnGlyGlyLeuCysAlaAlaLeuLysGluGlyCysCysPheTyrValAs 557
7455 GAGGAGAGACTGTGTGTGCTCCCAAAAGAGATGTGTTCCTAAGCGTA 7504
557 pHISerGlyAlaAlaLeuArgSerMetAsnLysLeuArgLysLysLeuG 574
7505 CCATACAGGCTGATGAAGATAGATATGGCCAAATTAAGAGAGACACTCT 7554
574 luarGATGATGATGluArgGluAlaAspGlnGlyTyrPheGlnGlyTyr 590


```

159 aIAsnSerGlyProGlyLysTyrLysMetLysLeuTyrLysasp... 174
1944 TAAACAACAATCTCACCTGTGACGAGCTGTGACAGTATGCAAAAGATAT 1993
175 LysSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrG1 191
1994 AAGGTGGTCAACCCC.....TTAGTTATTCGGTTTACAGA 2028
191 u.....ArgLysThrGlyLysTyrSerLysValAspLysTyrGluL 206
2029 CGCGGGAGAGCGGGTTACTTCTCGAGCACAGACATTTACTG..... 2071
206 euGlyAsnSerPheLeuLeuTyrGlyGlyAlaGlySerThrLeuThr 222
2072 ..GGCTTACGTTTGTATGTCTCGACAGAAGATCCAGGGCTTACATTGGG 2119
223 IleArgLeuArgTLeuThrGlyThrGluProProValAlaMetGlyPyr 239
2120 ATCCGACTCAGATACCA...AATCTAGAGACCCGCGTCCCAATVAGGCC 2166
239 oAspLysValLeuAlaGluGlnGlyProProAlaLeuGluProProHisA 256
2167 AAACCCCGTTCTGGAGACCAACAGCA.....CTCTCCAAGCCCAAA. 2209
256 snLeuProValProGlnLeuThrSerLeuArgProAspIleThrGlnPro 272
2210 ....CCTGTTAAGTCG.....CCTTCAGTCAACCAAAACA 2239
273 ProSerAsnSerThrThrGlyLeuIleProThrAsnThrProAlaGlnSe 289
2240 CCAGAGGG.....ACTCTCTCTCCCTACCCCACTTCCA..... 2275
289 rProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGln 306
2276 .....CCGGCGGAGACGAAATAGGCTGTAACTTACTAGACG 2315
306 lValAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCys 322
2316 GAGCCTACCAAGCCTCAACCTGACACAGTCTGACAAACCAAGAGTGC 2365
323 TrpLeuCysIleSerSerGlyProProTyrTyrGlnGlyMetAlaLysG1 339
2366 TGGTGTGTCTAGTAGCGGAGCCCTACTACTACAGAGGGTTCGCTCT 2415
339 uArgLysPheAsnValThrLysGlnHisArgAsnGlnCysThrTrpLys 356
2416 GGGTACCTACTCCAACCATACCTCTGCTCCAGCCACAGTCTCGTGGCT 2465
356 eArgAsnLysLeuThrLeuThrGlnValIleSerGlyLysGlyThrCysIle 372
2466 CCCAACACAAAGTTGACCTGTCCGAAAGTACCGGAGGAGGAGCTTCGATA 2515
373 GlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTy 389
2516 GGAGCAGTTCCCAAAACACATCAGCCCTATGATATACCAACCAAGACAAG 2565
389 rGlnGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpT 406
2566 CAGTCGAGGGTCC.....TATATCTAGTTGCCCTTACAGGTATACCATGT 2609
406 rPalAcysAsnThrGlyLeuThrProCysValSerThrSerValPheAsn 422
2610 GGGCTGTAGTACCGGGCTTACTTCATGCTCTCCACACCAACATAGTGAAC 2659
423 GlnSerLysAspPheCysValMetValGlnIleValProArgValTyrTy 439
2660 CTTACCACTCATATGTTGTCTGTGCACTCTGGCCAGAGTCACTCACTA 2709
439 rHisProGlnGluValValLeuAspGluTyrArgTyrArgTyrAsnArgP 456
2710 TCATTCCTCCCAAGCTATGTTTACGGCTCTTTGAG...AGATCCAAACCGAC 2756
456 rOlyArgGluProValSerLeuThrLeuAlaValMetLeuGly..... 470
2757 ACAAAAGAGAACCGGTCTGTTACCTTGCCCTGCTATTTATTTGGTGGACTA 2806
471 .....LeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeu1 485
2807 ACCATGGGGGAATTCGCTGGAATAGGAACAGGAGTACTAGTCTATAT 2856
485 eThrGlyProGlnGlnLeuGlnLysGlyLeuGlyIleLeuHisAlaIam 502
2857 G...GCCACTCAGCAATTCAG.....CAGTCCAAAGCCGAG 2891
502 eThrGlnAspLeuArgAlaLeuLysGlnSerValSerAsnLeuGlnL 518
2892 TACAGATATCTCAGAGAGGTTGAAAATCATCTTACCTTAAGAAAG 2941
519 SerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAs 535
2942 TCTCTCACTCCCTGCTGAAGTTGCTTACAGATCAAGAGGGGCTTGA 2991
535 pleuLeuPheLeuArgGlnGlyLysCysAlaAlaLeuLysGlnGluC 552
2992 GTTGTATTCTTAAGAAAGAGGCTGTGTCTCTTAAAGAAAGAT 3041
552 yscYsPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLys 568
3042 GTTGCTTATAGCGGACCAACAGACACTAGTACAGACAGATGGCCAA 3091
569 LeuArgLysLysLeuGlnArgArgArgArgGluArgGluAlaAspGlnG1 585
3092 TTGAGAGAGAGGCTTAATCAGAGACAGAACTGTTTGAGTCACTCAAGG 3141
585 yTrpPheGlnGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuS 602
3142 ATGGTTTGAGGACGCTTTAACAGATCCCTTGGTTTACCACTTGATAT 3191
602 eAlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGly 618
3192 CTACCATATATGGACCCCTCATTTGACTCTTATGATTTTCTTGGCA 3241
619 ProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAl 635
3242 CCTGCATTTCTTAATTCATTTAGTTCAATTTGTAAAGACAGATCTCAGT 3291
635 aValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 648
3292 AGTCCAGGCTTAGTCTGACTCAACATACCAACAGCTA 3331

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 02:58:34 ; Search time 212.72 Seconds

(without alignments)
3147.184 Million cell updates/sec

Title: US-09-171-553B-9

Perfect score: 2956

Sequence: 1 tgcctttaggttaggaac.....aaaaaaaaaaaaaaaaaaaaa 2956

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfillseq1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2889.4	97.7	3482	3	US-09-111-085-3
2	1882.8	63.7	7333	4	US-08-766-528-2
3	1563.4	52.9	2462	3	US-09-111-085-1
4	1489.2	50.4	8132	4	US-08-766-528-3
5	1231.8	41.7	8060	4	US-08-766-528-1
6	973.8	32.9	2000	4	US-09-376-781-25
7	973.8	32.9	2000	4	US-09-376-781-30
8	710.4	24.0	1493	4	US-09-376-781-24
9	455	15.4	6312	4	US-09-011-745-7
10	451.2	15.3	8535	3	US-08-716-351A-1
11	372.2	12.6	8655	3	US-09-075-272-1
12	357.8	12.1	10970	3	US-08-716-351A-5
13	343	11.6	638	4	US-09-376-781-18
14	307	10.4	6028	4	US-09-011-745-5
15	303.8	10.3	1965	1	US-08-258-420-9
16	301.4	10.2	2499	1	US-08-105-483-310
17	301.4	10.2	2499	1	US-08-709-209-310
18	301.4	10.2	2499	1	US-08-458-101-310
19	300.6	10.2	2001	3	US-08-850-961-13
20	288.8	10.1	8323	1	US-08-110-300A-8
21	288.8	10.1	8323	1	US-08-886-642-8
22	298.8	10.1	8323	5	PCT-US93-08041-8
23	298.8	10.1	1911	1	US-08-258-420-8
24	298.2	10.1	8202	1	US-08-552-369-13
25	296.8	10.0	1979	4	US-09-011-745-9
26	293.4	9.9	3925	4	US-08-552-369-19
27	290.2	9.8	1914	1	US-08-258-420-7

28	290	9.8	6061	4	US-09-011-745-6	Sequence 6, Appli
29	283.6	9.6	8332	3	US-08-850-961-1	Sequence 1, Appli
30	269.6	9.1	276	4	US-09-376-781-8	Sequence 8, Appli
31	253.4	8.6	278	4	US-09-376-781-10	Sequence 10, Appli
32	192	6.5	5109	1	US-08-073-836-1	Sequence 1, Appli
33	192	6.5	5176	1	US-08-073-836-2	Sequence 2, Appli
34	157.2	5.3	10306	3	US-08-716-351A-4	Sequence 4, Appli
35	144.4	4.9	9661	4	US-08-716-351A-3	Sequence 3, Appli
36	128.2	4.3	445	4	US-09-376-781-7	Sequence 7, Appli
37	113	3.8	10367	2	US-08-110-300A-9	Sequence 9, Appli
38	113	3.8	10367	2	US-08-886-642-9	Sequence 9, Appli
39	113	3.8	10367	5	PCT-US93-08041-9	Sequence 9, Appli
40	108.2	3.7	1237	1	US-08-798-000-2	Sequence 2, Appli
41	105.8	3.6	1374	1	US-08-278-630A-9	Sequence 9, Appli
42	103.8	3.5	1356	2	US-08-484-126-4	Sequence 4, Appli
43	102.8	3.5	735	1	US-08-798-000-3	Sequence 3, Appli
44	102.8	3.5	1234	1	US-08-798-000-1	Sequence 1, Appli
45	102.8	3.5	7308	4	US-09-011-745-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1	US-09-111-085-3
Sequence 3, Application US/09111085	
Patent No. 6100034	
GENERAL INFORMATION:	
APPLICANT: Sloye, Jonathan P	
APPLICANT: Weiss, Robin A	
TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope	
TITLE OF INVENTION: Specific sequences	
FILE REFERENCE: 4238/75168	
CURRENT APPLICATION NUMBER: US/09/111,085	
CURRENT FILING DATE: 1998-07-07	
EARLIER APPLICATION NUMBER: GB 9710154.7	
EARLIER FILING DATE: 1997-05-16	
NUMBER OF SEQ ID NOS: 16	
SOFTWARE: PatentIn Ver. 2.0	
SEQ ID NO 3	
LENGTH: 3482	
TYPE: DNA	
ORGANISM: Porcine retrovirus	
US-09-111-085-3	
Query Match	97.7%; Score 2889.4; DB 3; Length 3482;
Best Local Similarity	99.3%; Pred. No. 0;
Matches 2913; Conservative 0; Mismatches 21; Indels 1; Gaps 1;	
1 tgcctttaggttaggaacaccccttgacagtttggtgacccctatgaattgctt	60
548 tgcctttaggttaggaacaccccttgacagtttggtgacccctatgaattgctt	607
61 acgggggagccccccttgtagaattgcttctgacatagtgatgacgttctt	120
608 acgggggagccccccttgtagaattgcttctgacatagtgatgacgttctt	667
121 cccagccttgcttctcagagcgaagcgctcagtggtgtaggaacagcgctggaagc	180
668 cccagccttgcttctcagagcgaagcgctcagtggtgtaggaacagcgctggaagc	727
181 agctcggagagcctcaggaagagagagcttgaagtccatctgcttccaagt	240
728 agctcggagagcctcaggaagagagagcttgaagtccatctgcttccaagt	787
241 gagatcagctatgttagagcgcacgcgttcaggaacacgcagactcgtaggaagc	300
788 gagatcagctatgttagagcgcacgcgttcaggaacacgcagactcgtaggaagc	847
301 ctatctcgaactttgacacacacagcgctgtagaattcgaaggaattccacttga	360
848 ctatctcgaactttgacacacacagcgctgtagaattcgaaggaattccacttga	907

QY 361 tccatgcatcccaagcttaagctgagcgccacccctccgactcggggtgagagccgaaaga 420
|||||
Db 908 tccatgcatcccaagcttaagctgagcgccacccctccgactcggggtgagagccgaaaga 967
QY 421 ctgagaatcccttaagctgctccatcgccctgttcccttactctaaacataatcc 480
|||||
Db 968 ctgagaatcccttaagctgctccatcgccctgttcccttactctaaacataatcc 1027
QY 481 caagcagatgataagccttaagaacgctgaaccccaatagaaccttlaaccctaac 540
|||||
Db 1028 caagcagatgataagccttaagaacgctgaaccccaatagaaccttlaaccctaac 1087
QY 541 tggctgattatgacccctgataagcggtgtcactgttaaatagcaactcgaggtgttcc 600
|||||
Db 1088 tggctgattatgacccctgataagcggtgtcactgttaaatagcaactcgaggtgttcc 1147
QY 601 agagggacccctggtggtcgaactgcaattctgctccgattgattaaaccccgcttaaa 660
|||||
Db 1148 agagggacccctggtggtcgaactgcaattctgctccgattgattaaaccccgcttaaa 1207
QY 661 agcaacccctcccaacccctagtcggtatgaggttctatgtctgcccagcgacagagaa 720
|||||
Db 1208 agcaacccctcccaacccctagtcggtatgaggttctatgtctgcccagcgacagagaa 1267
QY 721 gagaataactggtggtgtcttgaggaactctctgtatagagatgaggtcgctcaactcc 780
|||||
Db 1268 gagaataactggtggtgtcttgaggaactctctgtatagagatgaggtcgctcaactcc 1327
QY 781 aagcatgagagactggaataatgacgactctctccagagacgggtlaaattctcctgttc 840
|||||
Db 1328 aagcatgagagactggaataatgacgactctctccagagacgggtlaaattctcctgttc 1387
QY 841 aatccggcccgaggcaagtaaaaaatgataaataataaagataaggtcgctcccca 900
|||||
Db 1388 aatccggcccgaggcaagtaaaaaatgataaataataaagataaggtcgctcccca 1447
QY 901 tcaactatgattatctaaagaataagtttcaactgaaaaaagaaacagaaataatctaa 960
|||||
Db 1448 tcaactatgattatctaaagaataagtttcaactgaaaaaagaaacagaaataatctaa 1507
QY 961 aagtgataaactggtatgagctggtggaataagtttttaataataatgagcggtgagagag 1020
|||||
Db 1508 aagtgataaactggtatgagctggtggaataagtttttaataataatgagcggtgagagag 1567
QY 1021 tccactttaaaccatctgctttagatagagagcggtgacaaacccctgtggaatgggga 1080
|||||
Db 1568 tccactttaaaccatctgctttagatagagagcggtgacaaacccctgtggaatgggga 1627
QY 1081 cccgataaagtaactggtcgaacaggggccccgggcccctgagcccaactaaactgccc 1140
|||||
Db 1628 cccgataaagtaactggtcgaacaggggccccgggcccctgagcccaactaaactgccc 1687
QY 1141 gtcgcccaattaaactcgctggtggtcgtgacataacacagccgctgcaacagtaaccac 1200
|||||
Db 1688 gtcgcccaattaaactcgctggtggtcgtgacataacacagccgctgcaacagtaaccac 1747
QY 1201 ggaattgattcttaacccaacgcttagaagaactcccaaggtcttccgttlaagaaagagag 1260
|||||
Db 1748 ggaattgattcttaacccaacgcttagaagaactcccaaggtcttccgttlaagaaagagag 1807
QY 1261 agactcttcagtctcatccagggagcttccaaagccatcaactccacgacccgtatgccc 1320
|||||
Db 1808 agactcttcagtctcatccagggagcttccaaagccatcaactccacgacccgtatgccc 1867
QY 1321 actcttctgttggcttgtgtatcctcctcaagcgctcctcttatttgagggagatgctaaa 1380
|||||
Db 1868 actcttctgttggcttgtgtatcctcctcaagcgctcctcttatttgagggagatgctaaa 1927
QY 1381 gaaagaaatctcaatgtgacccaagagatagaataatcaatgtatagtggttccggaat 1440
|||||
Db 1928 gaaagaaatctcaatgtgacccaagagatagaataatcaatgtatagtggttccggaat 1987

QY 1441 aagcttaccctcaactgaaattccgggaaggagcatgcataggaanaagctcccccaccc 1500
|||||
Db 1988 aagcttaccctcaactgaaattccgggaaggagcatgcataggaanaagctcccccaccc 2047
QY 1501 caaccaacaccttctgataagctgtgtgttattatgagcggtcctcaagaaatcaatctta 1560
|||||
Db 2048 caaccaacaccttctgataagctgtgtgttattatgagcggtcctcaagaaatcaatctta 2107
QY 1561 gtaacctgttataaagcggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1620
|||||
Db 2108 gtaacctgttataaagcggtggtggtggtggtggtggtggtggtggtggtggtggtggt 2167
QY 1621 tcaacttcaacaaatcccaagaattgtgtgtatgctggtccaaatcgctcccgaggttac 1680
|||||
Db 2168 tcaacttcaacaaatcccaagaattgtgtgtatgctggtccaaatcgctcccgaggttac 2227
QY 1681 taaccaatcggaggaatgt 1740
|||||
Db 2228 taaccaatcggaggaatgt 2287
QY 1741 gaaacccgtatcccttacccttagctgtatgctcgtatgagggagcggtgtgctaga 1800
|||||
Db 2288 gaaacccgtatcccttacccttagctgtatgctcgtatgagggagcggtgtgctaga 2347
QY 1801 acagggagagctgtccctgatacagaagccacagcgctagagaaagagcctgtgtagcta 1860
|||||
Db 2348 acagggagagctgtccctgatacagaagccacagcgctagagaaagagcctgtgtagcta 2407
QY 1861 catggtggtcctgacagaagaatctccgagccttaaaaggaatcgtgttagaaccctagaagag 1920
|||||
Db 2408 catggtggtcctgacagaagaatctccgagccttaaaaggaatcgtgttagaaccctagaagag 2467
QY 1921 tccctgactctctgtctgtaagtgtgtctacagaacccggagggatagatctgtgttt 1980
|||||
Db 2468 tccctgactctctgtctgtaagtgtgtctacagaacccggagggatagatctgtgttt 2527
QY 1981 ctaagagaaggt 2040
|||||
Db 2528 ctaagagaaggt 2587
QY 2041 tcaagagacatcagagacatccatgaaacaaagcttagaanaaagtttagagagcgctcgaag 2100
|||||
Db 2588 tcaagagacatcagagacatccatgaaacaaagcttagaanaaagtttagagagcgctcgaag 2647
QY 2101 gaaagagaggtgtgacccaggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2160
|||||
Db 2648 gaaagagaggtgtgacccaggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2707
QY 2161 accctgtcttctgtctcgaaggggccccctagtagctcctcctgttactcaagtggg 2220
|||||
Db 2708 accctgtcttctgtctcgaaggggccccctagtagctcctcctgttactcaagtggg 2767
QY 2221 cctgtcttaataatagttgt 2280
|||||
Db 2768 cctgtcttaataatagttgt 2827
QY 2281 atgtgtactttagcaacaggttaccagggcctctctgagccaagaaagaaactgacctagacc 2340
|||||
Db 2828 atgtgtactttagcaacaggttaccagggcctctctgagccaagaaagaaactgacctagacc 2887
QY 2341 tcccgattctaaagttagaactttaaacaagaagaagtgtgggaatgaaaggaatgaana 2400
|||||
Db 2888 tcccgattctaaagttagaactttaaacaagaagaagtgtgggaatgaaaggaatgaana 2947
QY 2401 tgcacactaaacccctcccaagaaacccaagaagttaataaagccttaaatgtcccccgaat 2460
|||||
Db 2948 tgcacactaaacccctcccaagaaacccaagaagttaataaagccttaaatgtcccccgaat 3007
QY 2461 acagacccctgt 2520
|||||
Db 3008 ccagacccctgt 3067
QY 2521 ctgtatacctgtgctaaagtaagataacagaagaatgagttgactaatcgcttatctgtgatt 2580

```

|||||
Db 3068 ctgcatctccgccaagtaagtaacagaaatggttgatactatctgcatc 3127
OY 2581 ctgtaaactactgagcacaatagaagaatgtatcacattgaagccctagtgacta 2640
OY 3128 ctgtaaacacgactgacacataagaaatgtatcacattgaagccctagtgacta 3187
OY 2641 tctcaactgcaatctgcaactctgccaagagagccacgacagatgagagactccgagacta 2700
Db 3188 tctcaactgcaatctgcaactctgccaagagagccacgacagatgagagactccgagacta 3247
OY 2701 tttaaatatgattgtgcacagagcgagcgctcgcgataatttaaatatgattgtgcacag 2760
Db 3248 tttaaatatgattgtgcacagagcgagcgctcgcgataatttaaatatgattgtgcacag 3307
OY 2761 gagcgagcgctc-tcgatatttaaatatgattgtgttgagcgacagcgctgtgtga 2819
Db 3308 gagcgagcgctcctcgatatttaaatatgattgtgttgagcgacagcgctgtgtga 3367
OY 2820 acccacaataaagctgtccgacttcgcgactcgagcgagcgagctcctcactccctcgctgtg 2879
Db 3368 acccacaataaagctgtccgacttcgcgactcgagcgagcgagctcctcactccctcgctgtg 3427
OY 2880 gtaacgactgtgccccagcgcgctgtggaataaaatcctctgtgtgtgtgcatc 2934
Db 3428 gtaacgactgtgccccagcgcgctgtggaataaaatcctctgtgtgtgtgcatc 3482

```

RESULT 2

US-08-766-528-2
Sequence 2, Application US/08766528

Patent No. 6190861

GENERAL INFORMATION:

APPLICANT: Jay A. Fishman

TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHAYE & COCKFIELD, LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIA TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766,528

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/572,645

FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers

REGISTRATION NUMBER: 35,965

REFERENCE/DOCKET NUMBER: MGP-038CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 7333 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-766-528-2

Query Match

63.7% Score 1882.8; DB 4; Length 7333;

Best Local Similarity 80.1%; Pred. No. 0; Mismatches 402; Indels 195; Gaps 9;
Matches 2409; Conservative 0;

```

OY 1 tgcctttaagtgatgaagaaacacccctgagcaagttgggtcagaccocctatgaattgctc 60
Db 4452 Tgcctttaagtgatgaagaaacacccctgagcaagttgggtcagaccocctatgaattgctc 4521
OY 61 acgaggagaccccccgcttgtagaattgctctgtatcatgctgtatgtctgtctt 120
Db 4522 ACGGGGGACCCCCCGCTTGAGAAATTCCTTTCACATGCTGTGATGCTGTGCTGCTT 4581
OY 121 cccagccttctctcctagctcgaagcgctgagtgagtgagtgagtgagtgagtgagtgag 180
Db 4582 CCCAGCCTTCTCTCCTAGCTCGAAGGCGCTGAGTGGTGGAGCGAGCGAGCGTGAAGC 4641
OY 181 agctccgagagcgctacacagagagagagagagagagagagagagagagagagagag 240
Db 4642 AGCTCCGGAGGCGCTCTACAGAG---GAGACTTGCAAGTTCACATGCTGTGCTGCTG 4698
OY 241 gagatcagctcctatgtagacgacgctgagagagagagagagagagagagagagagag 300
Db 4699 GAGATTCAGTCTATGTTAGAGCGACCGCTGAGAGAACTCGAGACTCGGTGAGAGGAGC 4758
OY 301 ctatctcgtactcttgaacacacacacacacacacacacacacacacacacacacacac 360
Db 4759 CTTATCTGTACTTTTACACACACACACACACACACACACACACACACACACACACAC 4803
OY 361 tccatgcatccacgtaagctgagcgacacccctccgactcgggtgtgagagcgagaaaga 420
Db 4804 ----- 4803
OY 421 ctgagatccccccttaagctcgtcctcagctgctgtgtcttactcctaacaataatccc 480
Db 4804 -----GATCCCTTAAGCTGCGCTCCATGCGGGGCTCTTCTGTGCAATAACTCCT 4859
OY 481 caagcgagtaagaacgccttaagaagcagcagcagcagcagcagcagcagcagcagcag 540
Db 4860 CAAGTTAAGTGAACGCTTGTGACAGCGCCGAACTCCCATTAACCTTATCTCTCACG 4919
OY 541 tgcgtgattatgaccctgatacaggtgtcactgtaaatgaacactcagatgtgtcctc 600
Db 4920 TGGTTACTTACTGACTCCGCTACAGGTATTAATTAATTAACACACACTCAAGGGAGGCTCC 4979
OY 601 aagagcaccgtgtgagctgacatgcttctcctcagatgattgaaccgcctgtta-- 658
Db 4980 TTGGGACCTGTGGCTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5039
OY 659 -----aagcagacccctcagctagctcgtgagctgagctgagctgagctgagctgag 711
Db 5040 GACCGAGCCACACCCCGATGACTCGTGTGCTTACGGCTTACGTTTACGTTGCGCAGACC 5099
OY 712 acagagaaagaaatactctgaggtgtctgaggaatcctctgtgagagatgagctgc 771
Db 5100 CCAATTAATGAAGAAATTTGGAATTCCTCAGGATTTCTTTTGCAAGCAATGAGAGCTGC 5159
OY 772 gtacactccaagatgagagacgtgaaatgagcagctcctcctcagagacgggttaaatc 831
Db 5160 ATAACCTCTAATGATGGAATTTGGAATTTGGAATTTGGAATTTGGAATTTGGAATTT 842
OY 832 tccctgtg-----caa 842
Db 5220 TCTTTTGTTAACAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 5279
OY 843 ttcgagccgagcaagtaacaatgtaacataataagaagaagctgtcccatc 902
Db 5280 GATTGGCAACGCGGTTACAAAGATG--TAGGAATTAAGCAAAATTAATTAATTAATTA 938
OY 903 agactgattatctaaaglaagttcactgtaaaaggaagaaacagaaatattcaaaa 962
Db 5339 AGACTGATTACTTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5398
OY 963 gtggataaagtgtgagctgaggaatgatttla--ataatgagcgagggagcgag 1019

```

Db 5399 GTGGTAAATGATATATCTTGGGGAATAGTACTATGAGGCTCTGGAGAAAGAAAG 5458
Qy 1020 gtcaacttaaccatctgccttagatagagcgggagacagaccctgtggcaatggg 1079
Db 5459 ATCTGTTCTGACTATGCTGCTCAGATAGAAATCAAGATGAGACCTCGGTTGCTATAGG 5518
Qy 1080 accgataaagtaactgtcgtgaacagggcccccgccttgagaccacgaactatgc 1139
Db 5519 ACCAAATAGGTTTGGCCGAACAAAGACCTCAATCCAAAGAAAAA----- 5560
Qy 1140 gggccccaattacactcgtcgtcgccgacatacaacagccgcttagcaacagtaacc 1199
Db 5561 -----ACAGAGGCCATCTCTTACCCCTCTGATTTACATTAACAACCTC 5602
Qy 1200 tggatgatctctaccacaacgccttagaactcccaaggtgtccgttaagaacagaca 1259
Db 5603 TGGATGAGTCCCACTGA-----GCCCTAACATCTATTAAACAGAGGCGC 5647
Qy 1260 gggactctcaactcactcagggagcttccagacacacacacacacacacacacacac 5647
Db 5648 GAAATCTTTTATGCTCATCCAGGAGCTTTTCAAGCTTAACTCCACGACTCCAGAGGC 5707
Qy 1320 caactctctgt 5707
Db 5708 TACCTCTTCTGTTGGCTTGTCTTACCTTCGGGCCACCTTACTATAGAGGATGGCTAG 5767
Qy 1380 agaaagaaatcaatgtgacacaaagagacagacacacacacacacacacacacacac 5767
Db 5768 AGAGAGAAATTCATATGTCACAAAGACATAGACACATATGATACATAGGAGATGCCAAA 1439
Qy 1440 taagcttaacctcaactgaagttccgggaagggagacatagagaaagaccccccac 1499
Db 5828 TAAGCTTACCTTACTGAGGTTTCTGAAAGAGGACCTGATAGGAGTGTCCCTCCATC 1587
Qy 1500 ccaaccaaaccttgcctactgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1559
Db 5888 CCACCAACACCTTTGTAACCACTGTAAGGCTTTATGTAACCTGTGAGAGTCAATATCT 5947
Qy 1560 agtaactgt 5947
Db 5948 GGTACTGCTGTATGACAGGAGTGGGATGATGATGATGATGATGATGATGATGATGATG 1619
Qy 1620 ctcaactcacaac 1607
Db 6008 CTGTGTTTCAACCAACTAAAGACTTGTGCTTATGCTTATGCTTATGCTTATGCTTATG 1679
Qy 1680 ctac 1607
Db 6068 CTACTATCCGAAAAAGCAGTCTGATGATGATGATGATGATGATGATGATGATGATG 1739
Qy 1740 agaacccgatacccttaaccacacacacacacacacacacacacacacacacacacac 16127
Db 6128 AGAGCCCATATCTCTGACACTAGCTGATGATGATGATGATGATGATGATGATGATG 1799
Qy 1800 aacagggagac 6187
Db 6188 AACAGAGAGGCTGCTCCCTTATACAGAGCGCAACAGCTGGAGAAAGAGACTTAACTACT 1859
Qy 1860 acatgvgcgaatgacagaaatctcagacacacacacacacacacacacacacacacac 6247
Db 6248 ACATCGAATGTTAAAGGAGATCTCAAGACCTTAAAGAAATCTGTAGTAACTGGAGGA 6307
Qy 1920 gtccctcaactctctgtgaagtggttcaagaacacacacacacacacacacacacac 6307
Db 6308 ATCCCTAACCTCTTATCTGAAGTGTCTTACAGAAAGAGAGGCTTAACTGTTATTT 6367
Qy 1980 tctaaagaaagtgggttctgtcagccttaaaagaaagtgtgtctctatlaagata 2039
Db 6368 TCTAAAGAGAGGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6427
Qy 2040 ctcaagagac 6427
Db 6428 CTCAGAGAGCATCAGAGACTCCATGAGCAAGCTTAAAGAAAGGTTAAGAGAGGCTGCAAG 6487

Qy 2100 ggaagagagcctgacacaggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2159
Db 6488 GGAAGAGAGGCTGACACAGGCTGTTTGAAGAGATGTTTAAACAGGCTCTTGTGATGAC 6547
Qy 2160 caacctgtctctgtcctgacggggcccttagtagtctcctgtctctctctctctctct 2219
Db 6548 CACCTGCTTCTGCTCTGACGGGCCCCAGTAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 6607
Qy 2220 gcttgccttaataagttgt 6607
Db 6608 GCTTGTCTTAAATATAGT 6667
Qy 2280 catgtacttaggac 2339
Db 6668 CATGTAATTAAGGACACAGTACCAAGGCTTCTGAGCCCAAGAGAACTGACCTTAGCC 6727
Qy 2340 ttcccaacttaagatlagactatlaaacaagaagtggtgtgtgtgtgtgtgtgtgtgtgt 6727
Db 6728 TTCCCACTTCTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 6787
Qy 2400 atgcaacttaaacctccac 6847
Db 6788 ATGCAACTTAACCTCCAGAACCCAGAACCCAGAACCCAGAACCCAGAACCCAGAAC 6847
Qy 2460 taacagac 6847
Db 6848 TCCAGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6907
Qy 2520 cctgtctatcctgtgac 6907
Db 6908 CCTGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6967
Qy 2580 tctgttaaacactgac 6967
Db 6968 TCTGTAAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 7027
Qy 2640 atccaactgcaactgtcactcctcaccagacacacacacacacacacacacacacac 7027
Db 7028 ATCTCAACTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2699
Qy 2700 attttaaatgtatgt 7087
Db 7088 ATTTTAAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2759
Qy 2760 ggaagcggggtcctgataattttaaatgtatgtgtgtgtgtgtgtgtgtgtgtgtgt 7147
Db 7148 GAGCGCGGCTCTGATATTTTAAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2819
Qy 2820 acccaataaagctgtccgactcgcacacacacacacacacacacacacacacacacacac 7207
Db 7208 ACCCATTAAGCTGTCCGATTTCCCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 2879
Qy 2880 gtacgactgt 7267
Db 7268 GTACGACTGT 2939
Qy 2940 aaaaaa 2945
Db 7328 AAAAAA 7333

RESULT 3
US-09-111-085-1
Sequence 1, Application US/09111085
Patent No. 6100034
GENERAL INFORMATION:
APPLICANT: Scye, Jonathan P
TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
FILE REFERENCE: 4238/75168
CURRENT APPLICATION NUMBER: US/09/111.085

TELEPHONE: (617) 227-7400
TELETYPE: (617) 227-5941

Qy 838 gtcaa-----ttccggccggcggaagracaaaatgatgaactatataaagat 888

TELEPHONE: (617) 227-7400
TELETYPE: (617) 227-5941

QY	886	aagagctgctccccatcagacttagattatactaagaatgaagtcttcacgtaaaaaggaaa	945
Db	6154	CCCAAGTGTCTTCCTTCAGACCTAGATTACTTAAAAATAGTTTCTCAGTGAAAGAGAAA	6213
QY	946	caggaaatatccaaagtgagataaatcggtatgagctggggaaatagttttta---ataa	1002
Db	6214	CAGAATAATATCCCTAAATGGGTAATAGTGTATCTCTTGGGAATGGTATATTATCGAGGC	6273
QY	1003	tatggcggggagcagggtccaattaaaccattgccttaggatagacagcgggacagaa	1062
Db	6274	TGGGGTAACAACACAGGCTCCATCTTAACATAATTCGCCCTCAAAATAAACAGCTG--GAG	6330
QY	1063	cccctgtggcaatgggacccgataaagtactggctgaacagggggcccgccctggag	1122
Db	6331	CCCTCAATGGGTATAGGACAAATACGGTCTTAGGGGTCAAAGAGACCCCAACC-----	6385
QY	1123	ccaccgataacttgcgggtgcccaataaacctgcctgcggcctgacatacacagccg	1182
Db	6386	-----AAGGACCAGGACCA	6399
QY	1183	octagcaacagtaccactggattctcacaacacgcctgataaactcccagggtgtt	1242
Db	6400	TCCTCTAACATAACTCTGGATCAGACCCCACTGAGCTAACAGCA-----	6445
QY	1243	cctgttaagcacagacagagactcttcagctctaccaggaggttcccagcatcaac	1302
Db	6446	-CGACTAAATGGGGCAAACTTTTAGCCTCATCCAGGAGGTTTTCAAGCTCTTAAC	6504
QY	1303	tccaccgacctgatgccactctcttgttggcttgtctatcctcagagcctcttat	1362
Db	6505	TCCACGACTCCAGAGGCTACCTCTCTTGTGGCTATGCTTAGTCTCGGCCCCACCTTAC	6564
QY	1363	tatgagggtatgctaaagaagaataatcaatgtgaccaagagcatagaatacaatgt	1422
Db	6565	TATGAGGAATGGCTAGAGAGGGAAATTCATATGTGCAAAAGAAATAGAGACCAATGC	6624
QY	1423	acatgggggtcccgaaataagcttacccctcactgaagtttccgggaaggagcatgcata	1482
Db	6625	ACATGGGATCCCCAAATAAGCTTACCTTACTGAGGTTCTGGAAAGGCACCTGCATTA	6684
QY	1483	ggaaaagctcccccattcccacacaacactttgtatagctatgtttagcagcggcc	1542
Db	6685	GGAAGGTTCCCCCATCCCACCAACACCTTTGTAAACACACTGAAGCCTTTAATCAAAACC	6744
QY	1543	tcagaaaatcagtatatttagtacctggtataacaggttggtggcctgcaatactcgggta	1602
Db	6745	TCTGAGAGCAATATCTGTTACCTGGTTATGACAGGTGTGGGCATGTAAATACTGGATTA	6804
QY	1603	acccctgtgtttcacctcngtctcaacaaatccaaagatttgtgtcatggtccaa	1662
Db	6805	ACCCCTGTGTTTCCACCCTGGTTTTTAAACCAAACTAAAGATTTTGTGCATTTATGTCCAA	6864
QY	1663	atcgtcccccgagtgtaaccatcctcggaagatggtccttgatgaataatgactatcgg	1722
Db	6865	ATTGTTCCCGGAGTGTATATCTATCCCGAAAAAGCAATCCTTGATGAATATGACTTACAGA	6924
QY	1723	tataaccgacaaaagagaaccogtataccttaccctagctgtgaatgctcggaatagg	1782
Db	6925	AATCATCGAAAAAGAGAGAACCCATATCTCTGACACTTGTCTGTGATGCTCGGACTTGA	6984
QY	1783	acggccgttggcgtagggaacagggaagctgcctgtatcacaggaccacagcagctagag	1842
Db	6985	GTGGCAGAGGTGTAGBAACAGGAACAGCTGCCTGGTCACGGGACCAACAGCAGCTAGAA	7044
QY	1843	aaaggacttggtagctacatgcgcgcgcgaatgacagaagatctccgagccttaaaagagct	1902
Db	7045	ACAGGACTTAGTAACCTACATCGAATTGTAAACAGAAGATCTCCAAGCCCTAGAAAAATCT	7104
QY	1903	gttagcaactagaagagtccctgaacttttgtcgaagtgtttctacagaacccgagg	1962
Db	7105	GTCACTAACCTGGAGGAATCCCTAACCTCCTTATCTGAAGTAGTCTCTACGAATAGAGA	7164

QY	1963	ggattagatctgctgtttcttaadagaaaggtgggttatgtgcagccttaaaagaagaatgt	2022
Db	7165	GGGTTAGATTTTATTATTCTTAAAGAAGAGGAGGATTATGTGTAGCCTTGAGGAGGAATGC	7324
QY	2023	tgtctctatgtagatcaactcaggagccatcagagactcctgaacaagcttagaaaaag	2082
Db	7225	TGTTTTATGTGGATCATTTTCAGGGGCATCAGAGACTTCCATGAACAAGCTTTAGAGAAAGG	7284
QY	2083	ttagagaggcgtcgaaagggaaagagaggctgaccagggtgtgtttgaaggatggttcaac	2142
Db	7285	TTGAGAGAAGCTCGAAGGGAAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGTTCTCAAC	7344
QY	2143	aggctccttgtagtagcacccctgcttctcgtctgaaggggcccttagtagtctgctc	2202
Db	7345	AGGTCCTCTTGGTTGGCTACCCCTACTTCCTGCTTTTAAAGGACCTTTAATAGTCTCTCTC	7404
QY	2203	cgttacttacagttgggcttgccttaattaataggtttgttgccttttagagaaacga	2262
Db	7405	CTGTTACTCACAGTTGGGCATGCTATTATTAACAAGTTAATTGCTTCATTAGAGACGA	7464
QY	2263	gtgagtcagtcacagatcatgtacttaggcaacagtcaccaaggcctctgaccaagga	2322
Db	7465	ATAAGTCAGTCCAGATCATGTGTACTTTAGAACACAGTACCAAGGCCCGCTCTAGC---	7521
QY	2323	gaaactgaccttagctctccagttctaaattagaactattaaacagacaagaagtgg	2382
Db	7522	GAAGCTGGCCGCTTAGCTCTTACCAGTCTTAAGATTAGAACTATTAAACAAGAGAAGAGTGG	7581
QY	2383	ggaatgaagagatgaaatgcaacctaaacctccacagaccagggaagttaataaaaaagc	2442
Db	7582	GGAAATGMAAGGATGAATAATACAACTTAGCTTAATGAGAAGCTTAAATTTGTTCTGAATTC	7641
QY	2443	tctaaatgccccgaattacagaccctgctggctgccagtaataataggtagaaggtccac	2502
Db	7642	CAGAGATTGTGTTCTTATAGGTAAAAATATTAGCTTTTGTGTTTAAAAATATCGGAAG	7701
QY	2503	ttcctattgtccaggcctgctctatccctggcct-aagtaagaataacagggaagtgtga	2561
Db	7702	TAAAAATAGCCCTGAGTACATGCTCTTAGGCATGGAACCTTCTTGAACACTATTTTGAGATAA	7761
QY	2562	c----taatgcttactggattctgttaaactgactgcccaccatagaagaattgattac	2617
Db	7762	CAAGAAAAAGGAGTTCTTAACCTGTGTTTGTAGCTTCTGTAAAACTGGTTCGCCCATAAAG	7821
QY	2618	acattgacagccctagtgacctaatacactgcaatctgcaatctgcccagagcgcggtctcga	2677
Db	7822	ATGTTGAATGTTGATACACATATCTTGGTGACACATGCTCCCCCCCCCGGAAACATG	7881
QY	2678	gcagatgagcactccgagcattttaaatgattggtccagcgagcgcggtctctcga	2737
Db	7882	CGCAATGTGTAACTCTAAAAAATAATTAATTAATTTGGTCCAGAAAGCGCGGCTCTCGA	7941
QY	2738	tattttaaatgattggttccatggagcgcggtctctcgaattttaaatgattggtttg	2797
Db	7942	AGTTTTAAATTGACTGGTTCT-----GATATTTGAAAGTATGGTTGTG	7986
QY	2798	taacgcagagctttgttgyaaccccataaaaagctgtcccgcattccgcactcggggcgc	2857
Db	7987	TAAAGCGGGCTTTGTGTGTGAACCCCATAAAAAGCTGTCCCAGACTCCACACTCGGGGCG	8046
QY	2858	cagtccttacccctgcgtggtgtagactgtggcccccagcgcgcttggataaaaaatc	2917
Db	8047	CAGTCTCTACCCCTGGCTGGTGTACGACTGTGGGGCCCCAGCGGCTTTGGAAATAAATC	8106
QY	2918	ctcttgctgtttgcatcaaaaaaaa	2943
Db	8107	CTCTTGTGTTGTGATCAAAAAAAA	8132

RESULT 5
US-08-766-538-1
; Sequence 1, Application US/08766528
; Patent No. 6190861

Db 1429 TGCTGTGATGTCGACCTTGAGTGGCAGAGGTGTAGGAACAGGACAGCTGCCCTGCT 1488
Qy 1821 cacagaccacagcagctagagaaagacttggtgagctacatcgcgccatgacagaaga 1880
Db 1489 CACGGGACCACAGCAGCTAGAAAACAGACTTAGTAACCTACATCGAATTTGAACAGA 1548
Qy 1881 tctccgagcctaaagagctctgttagcaaccttagagagctccctgaactctttgtctga 1940
Db 1549 TCTCCAAGCCTAGAAAAATCTCTCAGTACCTGGAGGAATCCCTAACCTCTTATCTGA 1608
Qy 1941 agtggcttcacagaccgaggggattagatctgtcttctaaagagaggtgggtatg 2000
Db 1609 AGTAGTCTACAGAAATAGAGAGGGTTAGATTTATTTCTTAAAGAGGAGGATTATG 1668
Qy 2001 tgcagccttaaaagaagaatgttgctctctatagatacactcagagagccatcagagactc 2060
Db 1669 TGTAGCCTTGAGAGGAATGCTGTTTTATGTGGATCATTCAGGGGCCATCAGAGACTC 1728
Qy 2061 catgaacagcttagaaaaaagttagagagggcgctgaaggggaagagagggctgaccagg 2120
Db 1729 CATGAACAACTTAGAGAAAGGTGGAGAGGCTCGAAGGGAAGAAACTACTCAAGG 1788
Qy 2121 gtgtttgaagatggttcaacagctctccttggatgacacccctgtcttctgtctgac 2180
Db 1789 GTGTTTGAGGATGTTCAACAGGCTCCTTGGTTGGCTACCCCTACTTTCTGCTTAA 1848
Qy 2181 ggggcccctagtagtctgctctgttacttactacagttggccttgcttaataatagtt 2240
Db 1849 AGGACCTTTAAATAGTCTCTCTCTTACTCAGAGTTGGGCGCATGTATTATTAACAAGT 1908
Qy 2241 tgtgctttgttagaacagagtagtgctgctccagatcatggtacttaggcaacagta 2300
Db 1909 AATTGCCCTTCATTAGAGAACCAATTAAGTGCAGTCCAGATCATGTACTTAGACAACAGTA 1968
Qy 2301 ccaagggccttctgagccagagagaaactgacctctgacctccagcttccagcttcaagattagaa 2360
Db 1969 CCAAGGCCGCTAGC---AGGGAAGCTGGCCGCTAGCTCTACAGGTCTTAAGATTAGAA 2025
Qy 2361 ctatttaacagacagaagtgaggaaatgaagagatgaataatgaacacccctccacaga 2420
Db 2026 CTATTAAACAGAGAGAGAGTGGGAATGAAGGATGAATAACAACCTAAGCTAATGAGA 2085
Qy 2421 accaggaagttaataaaagctctctaaatgccccgaattacagacccctgctggctgcca 2480
Db 2086 AGCTTAAATTTGTTCTGAATTCAGAGTTTGTTCCTTATAGGTAAAGATTAGGTTTTT 2145
Qy 2481 gtaaataggtagaaggttcacacttcttcttctccagggcctgctatcctggcct-aagt 2539
Db 2146 CTGTTTAAATAATGCGGAAGTAAATAGGCCCTGAGTACATGTCTCTAGGCATGAAC 2205
Qy 2540 aagataacaggaatgagttgac---taatcgcttctctggtattctgttaaacactgctg 2595
Db 2206 TCTTGAACACTATTGTAGATAACAAGAAAGGGAGTTCTTAAGTCTTTTAGCTTCTG 2265
Qy 2596 gcaccatagaagaattgattacacattgacagccctgagctatctcaactgcaact 2655
Db 2266 TAAACCTGGTTGCGCCATAAGATGTGAAATGTTGATACACATATCTTGGTCAACAAT 2325
Qy 2656 gtcaactgcccagagaccacacagatcgagacccctcgagctatttttaaatgattgg 2715
Db 2326 GTCTCCCCACCCGAAACATGCGGAAATGTGTAACCTCTTAAACAATTTAAATTAATGG 2385
Qy 2716 tccagagagcgggctctcgatttttaaaatgattggttccatgagcgcggtctctg 2775
Db 2386 TCCACGAAGCGCGGCTCTCGAAGTTTTAAATGTACTGTTGT-----G 2430
Qy 2776 atatttaaatgattggtttgtgacgacaggtttgtgtgaaccccatataaagctgt 2835
Db 2431 ATATTGTGAATGATTGGTTGTGTAAGCGCGGGCTTTGCTGTGAACCCCATATAAGCTGT 2490
Qy 2836 ccgagattccgactcgggcgccagtcctctacccctcggttggtgacgactgtggccc 2895
Db 2491 CCGGATCCACACTCGGGGCCCGAGTCTCTACCCCTCGGTGGTGTAGACTGTGGGCC 2550

Qy 2896 cagcgccgttggataaaaaatctcttctgtgtttgcatcaaa 2938
Db 2551 CAGCGCGCTTGGATAAAAAATCTCTGTGTGTTCATCAAGA 2593

RESULT 6

US-09-376-781-25
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376.781
; EARLIER FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOS: 18, 22, and 24.
US-09-376-781-25

Query Match 32.9%; Score 973.8; DB 4; Length 2000;
Best Local Similarity 70.2%; Pred. No. 1.2e-305;
Matches 1461; Conservative 0; Mismatches 532; Indels 88; Gaps 8;
Qy 332 caccacacgctgtgaagtcgaaggaatctccacgtgattccatgcacccacgttaagc 381
Db 1 caccacacgctgtgaagtcgaaggaatctccacgtgattccatgcacccacgttaagc 60
Qy 382 tggcgccacctcccgactcggggtgagagccgaaagagactggaatcccttaagcttc 441
Db 61 cggcgccacctcccgactcggggtgagagccgaaagagactggaatcccttaagcttc 120
Qy 442 gctccatcgctgttcttcttacttacttaacataactcccgagccagtagtaaacgctt 501
Db 121 gctccatcgactgttcttcttacttctgtaataacacctctcagactaatggtatgcata 180
Qy 502 atagacagctcgaaaccccatagacctttatcccttacctgctgattattgacctgat 561
Db 181 ggaacagcctgaaaccccatagacctttatcccttacctgctgattattgacctgat 240
Qy 562 acgggtgtcactgtaaatagcactcgagggtgttgcctcctagagcagcctggtggcctgaa 621
Db 241 acaggtattatattccacagcgtcgagggaggtcctctttagaacctggtggcctgat 300
Qy 622 ctgacttctgctcgattgattaaaccccgctgtttaaagcacacacctcccaacctagtc 681
Db 301 ctatagctgctcgactgactgacttc---ctagctgacctcaacccagatatactc 357
Qy 682 cgtagttatgggtcttctgctgctcagg---cacagagaaagagaaatactgtgggggt 738
Db 358 cgtgcttgcgactgactgactgacttc---ctagctgacctcaacccagatatactc 417
Qy 739 tctgggaatcctctgttaggagatggagctgcctccactccacagtgagactggagactgaaa 798
Db 418 cctagagatttcttttacaacaaatggagctgtgtaacctctaatgatggaatcgaaa 477
Qy 799 tggcagatctctccagacgggtgaaatctcttctgttaattccggtggcggcgaag 858
Db 478 tggcgaacctctctcagagtagggtaagcttttcttatgtcaaccccaataaccaccgg 537

QY 859 tacaaaatgtagaaactataaagataagagctgtctcccatcagacttagattatcta 918
 Db 538 acctggaagacgtacacaggtaccattctgggttttccctcagacactgattatctt 597
 QY 919 aagataagttccactgaaagaaagaaagaaatattcaaaagtggaataatggtatg 978
 Db 598 aaaaataagttccacgaaagaaagaaagaaatattcaaaagtggaataatggtatg 657
 QY 979 agctgggaatagtttttta---ataatattggcggggagcagcaggttccactttaaccatt 1035
 Db 658 tcttggtgaaataatattataacaggttgagacagacacacaggttccactttaaccatt 717
 QY 1036 cgccttagtagagacgggggacagaaacccctgtggaataggagacccgataaagtaactg 1095
 Db 718 cgaactaaata---agcagctagagctcccaatggtctatagacagcaatacggctctta 774
 QY 1096 gctgaacagggggcccccggccctgagccacccgacataaacttgcgggtgcccataaacc 1155
 Db 775 agcgggtcaaaagaaaccccccagccagccatccctct-----810
 QY 1156 tcgctgcggcctgacataacacagcgcctagcaacagtagtaacacagtaggattgatttoactc 1215
 Db 811 -----gataaacttctaaattagacccact 837
 QY 1216 aacacgcctagaaactcccgaggtgtctcttaagacagacagacagactcttcagtcctc 1275
 Db 838 gagtctaacagcagcagc-----taaaaggggacaaaacttttttagctc 882
 QY 1276 atccaggagcttcccaagccatcaactccacccagccctgtgacactctctctgttgg 1335
 Db 883 atccaggagcttcccaagccatcaactccacccagccctgtgacactctctctgttgg 942
 QY 1336 ctttctctatctccagggcctctctattataggggagtgcttaagaaagaaagaaattcaat 1395
 Db 943 ctttctctatctccagggcctctctattataggggagtgcttaagaaagaaagaaattcaat 1002
 QY 1396 gtacacaaagagacagaaatacaatgtacatgggggtcccgaaataaagcttaccctact 1455
 Db 1003 gtacacaaagagacagaaatacaatgtacatgggggtcccgaaataaagcttaccctact 1062
 QY 1456 gaagtttccggggaaggggacatgcataagaaagctcccccacccacacacacactttgc 1515
 Db 1063 gaagtttccggggaaggggacatgcataagaaagctcccccacccacacacacactttgc 1122
 QY 1516 tatagtactgtttatagcagcgcctcagaaaaacagttatttagtactggtttataac 1575
 Db 1123 aaccacactgaagcctttaatacaacactctgagagccaatactggtacctggttatgac 1182
 QY 1576 aggtgtgggcatgcaatactgggttaaacccctgtgttccacactcagcttccaccaa 1635
 Db 1183 aggtgtgggcatgcaatactgggttaaacccctgtgttccacactcagcttccaccaa 1242
 QY 1636 tccaaagattgtgtcatgttccaaatcgtcccccaggtgtactaccactcgtgagaa 1695
 Db 1243 actaaagactttacattggtccaaatgttcccccaggtgtactaccactcgtgagaa 1302
 QY 1696 gtggtccttgatgaatgactatcggtataacccgacaaagagacccggtatccctt 1755
 Db 1303 acaattctcgatgaatgattacaggaacccatcgacaaagaaacccatcccttcc 1362
 QY 1756 accctagctgtaattgctcgattaggaagcggcgtgtggtaggaacagggagcagctgcc 1815
 Db 1363 acactcgcagtaattgctcgactcgagtgatgaacaggtgtggaacaggaactgcagct 1422
 QY 1816 ctgatacaggaacacagcagctgagaaagacgttggtagctacatcgccgacatgaca 1875
 Db 1423 ttagttaacaggaactcagcagctgagaaagacgtttagtaacctacacataaattgtaoa 1482
 QY 1876 gaagatctccagccttaaaagagctctttagcaacactagaagagctccctgactctttg 1935
 Db 1483 ggaaatctccagccttaaaagagctctttagcaacactagaagagctccctgactctttg 1542

QY 1936 tctgaagtggttctacagaaccggagggattagatctgtctttcttaagaaaggtggg 1995
 Db 1543 tctgaagtggttctacagaataaaaggggttagattttattttcttaaaagaaagaga 1602
 QY 1996 ttatgtcagccttaaaagaaagatgttgccttctatgtatgcactcagagccatcaga 2055
 Db 1603 ttatgttagccttaaaagaaagatgttgccttctatgtatgcactcagagccatcaga 1662
 QY 2056 gactccatgaacagccttagaaaaagtttagagagcgcgtcgaaagggaaagagagcgtgac 2115
 Db 1663 gactccatgaacagccttaaaaaagtttagagagcgcgtcgaaagggaaagagagcgtgac 1722
 QY 2116 -caggggtggtttgaagagatggttcaacaggtctcttgcagtcacacccctgctttctgc 2174
 Db 1723 tcaagatggttttaaaagatggttcaacaggtctcttgcagtcacacccctgctttctac 1782
 QY 2175 tctgaagggggcccttagtctctctcttcttacttacagttggccttgccttacttaa 2234
 Db 1783 ttaacaggaaccccttaagtagtccctcccttcttactcacagttggccttacttaaa 1842
 QY 2235 taggtttgtcccttttagaagacaggtgagtgagtgagtcagtcagtcagtcagtcagtc 2294
 Db 1843 caagtttaattgcttcttagaagacaggtgagtgagtgagtcagtcagtcagtcagtcagtc 1902
 QY 2295 acagtcacaaagccttgcagcaggaagaaactgactcttagccttccagttcttaaga 2354
 Db 1903 acagtcacaaagccttgcagcaggaaggtggggaatgaaaggt 2395
 QY 2355 ttagaactattaaacaaagaaaggtggggaatgaaaggt 2395
 Db 1960 ttaaaactattaaacaaagaaaggtggggaatgaaaggt 2000

RESULT 7
 US-09-376-781-30
 : Sequence 30, Application US/09376781
 : Patent No. 6261806
 : GENERAL INFORMATION:
 : APPLICANT: Banerjee, Papi T.
 : APPLICANT: Patience, Clive
 : APPLICANT: Andersson, Goran K.
 : TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
 : Patent No. 6261806
 : FILE REFERENCE: 61750-267
 : CURRENT APPLICATION NUMBER: US/09/376,781
 : CURRENT FILING DATE: 1999-08-18
 : EARLIER APPLICATION NUMBER: 60/097,015
 : EARLIER FILING DATE: 1998-08-18
 : NUMBER OF SEQ ID NOS: 33
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 30
 : LENGTH: 2000
 : TYPE: DNA
 : ORGANISM: Artificial Sequence
 : FEATURE:
 : OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
 : OTHER INFORMATION: consensus sequence of PERV-D.
 US-09-376-781-30

Query Match 32.9%; Score 973.8; DB 4; Length 2000;
 Best Local Similarity 70.2%; Pred. No. 1.2e-305;
 Matches 1461; Conservative 0; Mismatches 532; Indels 88; Gaps 8;

QY 322 caccacagcgtgtgaaagtcgaaggaatctccacccctggatccatccacgttgaagc 381
 Db 1 caccacagcgtgtgaaagtcgaaggaatctccacccctggatccatccacgttgaagc 60
 QY 382 tggcgcacccctccgactcgggtggagcgcgaagagctgagaatcccttaagcttc 441
 Db 61 cggcgccacccctccgactcgggtggagcgcgaagagctgagaatcccttaagcttc 120

QY	442	gctccatcgctgggttccttactactaaataaactcccagcgccagtagtaaacgctt	500
Db			
Db	121	gctccatcacatgggttccttactctgtcaataaactctcagactaatggatgycaca	180
QY			
QY	502	atagacagctcgaacccccatagacatttaccctacactgggtgatttaccctgat	561
Db			
Db	181	ggagacagctgaactcccataaaccttactctgactcgtttaattactgactctgac	240
QY			
QY	562	acgggtgctcaactgtaaaTAGCACTCGAGGTGTGCTCCTAGAGCAACCTGGTGGCTGAA	621
Db			
Db	241	acaggtattaatccacagcgtcgaggggagctctttagaaacactcgggtggctgat	300
QY			
QY	622	ctgcattctgcctcggattgattaaocccgcgtttaaaagcacactcccaactagtc	681
Db			
Db	301	ctatatgtcgtcctcagatcagtcattc---ctagtcgactccaacccaagatactc	357
QY			
QY	682	cgtagtgtggttctatgtgcceagg---cacagagaagagaaaactctgtgtggggt	738
Db			
Db	358	cgtgttcaggtatttatgtttgccaggaccaccaataatggaaaaactatggaat	417
QY			
QY	739	tctggggaaactctctgtagagatggagctgcgtcacctccaaagatggagactggaaa	798
Db			
Db	418	cctagagattctttacaaaacatggagctgtgaacctctaatgctggaatcggaaaa	477
QY			
QY	799	tggcgaatctctctccaggacccgggtataaattctcttgtcaattccggccggcgcaag	858
Db			
Db	478	tggccaaactctgcgagataggtaagcttttctttatgtcaacccaataaacacagg	537
QY			
QY	859	tacaaaatgataaactatataaagataagagctgtccctccatcagaactagattatcta	918
Db			
Db	538	acctggaaaacgtctacaggtaccattctggtgttttccctcagactagattatctt	597
QY			
QY	919	aagataagtttcaactgaaaaaggaaacacaggaaaaatttccaaaagtgataatgggatg	978
Db			
Db	598	aaaataagtttccacggaaaaaaaacaaagaaaatactctaaatggataaatggtatg	657
QY			
QY	979	agctggggaaatagtttita---ataatggcgggggagcagggtccactttaaccatt	1035
Db			
Db	658	tcttggggaaataataatatatacaggttgggacagacacacaggctccattctaaacct	717
QY			
QY	1036	cgcttaggataagacggggacagaaacccctgtgycaatgggacccgataaagtactg	1095
Db			
Db	718	cgacttaaaaa---agccagctagagctcccaatgggtctagagacgaatacgggtta	774
QY			
QY	1096	gctgaacaggggcccccgccctggagcgcgcgataaacttgcgggtgcaccaataacc	1155
Db			
Db	775	acgggtcaagaagaccacccccggacactcctc-----	810
QY			
QY	1156	tcgctgcgctgacataacacagcgcgctgacagtagtaccactggtgattgattctacc	1215
Db			
Db	811	-----gatataacttctaaattagacccact	837
QY			
QY	1216	aacacgcctagaaactccccagggttctcgtgttaagacaggacagagactcttcagtcct	1275
Db			
Db	838	gagctaaacagcaac-----taaaacggggacaaacttttagtctc	882
QY			
QY	1276	atccaggagctttccaaagccatacaactccaccgacctgatgccacttctcttgttgg	1335
Db			
Db	883	atccaggagcttttcaagctctcctaactccacgactccagaggctacctctcttgttgg	942
QY			
QY	1336	cttctctatcctcaggccctcttattatgagggatggctaaagaagaataatcaat	1395
Db			
Db	943	ctttgcttaactctcgccccacttactataagaataatggctaaaaagagaaaaatcca	1002
QY			
QY	1396	gtgaccaaagagcatagaaaatcaatgtacatgggggtccccgaataagcttacctcact	1455
Db			
Db	1003	gtgacaaaaaacatagagaccaatgtacatggggaTccccaaataagcttaccttact	1062
QY			
QY	1456	gaagtttccggaaagggacatgcatagggaaaagctccccctccccacacacactttgc	1515
Db			
Db	1063	gaggtttctggaaaggacactgcataaaaaaaggttccccctccccacacacactttac	1122
QY			
QY	1516	tatagctagtgtgtttatgacagggcctcagaaaaatcagtatattagtaactggtttataac	1575

RESULT 8

RESOLUTION 8
US-09-376-781-24

03 03 370 701 24
; Sequence 24, Application US/09376781

; Patent No. 6261806

; GENERAL INFORMATION:

APPLICANT: Banerjee, Papia T.

APPLICANT: Patience, Clive

; APPLICANT: Andersson, Goran K.

;
TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of

; Patent No. 6261806

; TITLE OF INVENTION: Use

; FILE REFERENCE: 61750-267

; CURRENT APPLICATION NUMBER: US/09/376,781


```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4058)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4059)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4060)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4061)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4246)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4247)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4248)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4249)
; OTHER INFORMATION: n is any nucleotide
; US-09-011-745-7

Query Match          15.4%; Score 455; DB 4; Length 6312;
Best Local Similarity 53.0%; Pred No. 8.5e-137;
Matches 1261; Conservative 0; Mismatches 1030; Indels 88; Gaps 10;

QY 2  gcttttagggtagaacacccctggacagtttgggctgacccctatgaattgctcta 61
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 1208 gctgcttagggccaggaaataccctggcgggttggtttaactctcttaagaattctcta 1267

QY 62  cgggggaccccccctggtagaaatgcttctgtacatagctgtgtgtgttctc 121
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 1268 tggaggacacccccatacttgtagctgtggagaaacttgggtcccgatagatttct 1327

QY 122  ccagccttgttcttaggctcaaggcgtcgagtggtgaggaacagcgtggaagca 181
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 1328 cccgttattactacttaaggcttaaggctttagaaattgagggacccaaatctgggacca 1387

QY 182  gctccggaggcctactcagagagagagacttgcaggttccacatcgtctccaagtgg 241
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 1388 gatcaagagggtgataagcctg---gtaccgtaaacatccctcaccgttccaggtcgg 1444

QY 242  aqattcagctctatgtagaccacccgtgcaggaaacccctcagactcgtgtgaaaggacc 301
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 1445 gsatcaagtgtgtcagacgcatcagaccagcagccttgagctcgtgtgaaaggccc 1504

QY 302  ttatctgtacttttgaccacacacagcgtgtgaaagtgcagaaagaaatccaccctggat 361
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 1505 atacctggtgtgtactaccgcagccggttaaaagctcgtgtgtgtgtgtgtgtgtgtgt 1564

QY 362  coatgcataccacgttaagctggggccacct-----cccgactcgggtgagagacgca 415
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 1565 ccattgtctcacctcaaacctgcaccacccctcggcaccagatgagtcctgggagcgtgga 1624

QY 416  aaagactagaaatcccttaagcttgcctccatcgcctgcttcttactct-aaacata 474
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
```


SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,272
FILING DATE: 08-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,140
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: POOF, Brian W.
REGISTRATION NUMBER: 32,928
REFERENCE/DOCKET NUMBER: 14538A-003710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 467-9600
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8655 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-075-272-1

Query Match 12.6%; Score 372.2; DB 3; Length 8655;
Best Local Similarity 62.1%; Pred. No. 8.3e-110;
Matches 663; Conservative 0; Mismatches 383; Indels 21; Gaps 4;

QY 1250 agacaggacagacgttcagtcctcagggagaggtttccagccatcaatccaccg 1309
DB 6706 ATACGAAACCGTGTGCTAGCTAGTTCAGGGAGCCCTTTTAAATAGGACTA 6765
QY 1310 accctgagccactctctgtgtgtgtgtctatccctcagggcctctattatgag 1369
DB 6766 ATCTTAATGACTCAATCAGTTGGTTATGCTATGCTCTTAACCCCTTATATGAAG 6825
QY 1370 ggaatggaagaagaataatcaatgacaaagagcatagaaataatgacatggg 1429
DB 6826 GAATTGCTCAGACTAGAACTTACATATATCTTACATGATCA--TCTCAATGCTTTGGG 6882
QY 1430 ggtccgaaataagcttccctcactgaagttccggagagggacatgcataagaa 1489
DB 6883 GAGAAACAGGAAATGTGACTCGACGAGTTTCAGGAATGGGCTTTGTTAGGTGAG 6942
QY 1490 ctcccccatcccaacacacctttgctatagctgtgtgtgtgtatgagcagcctcagaaa 1549
DB 6943 TGCCCCAGGATAAATGGCACCTCTGTACCCAGACTCAAAATATCCGACCTAACAAAGTG 7002
QY 1550 atcagattatgacctgtgtataacaggtgtgtgggcatgaactactgggttaaccccc 1609
DB 7003 GTCAATATCTAGTGTCTCCCATAGACAGATATGGGCTTGCATACAGGTCTCACTCCT 7062
QY 1610 gtgttccacctcagcttcccaacaaatccaaagattgtgtcattgtccaaatcgtcc 1669
DB 7063 GTATATCATGTCTGCTTTTCAATAGCTCCAAAGATTCTGTATTTAGTTCAGCTTATTC 7122
QY 1670 cccgagtgactaccatcctcaggaagtgtgtccttgatgaatgactatcggtataacc 1729
DB 7123 CTAGACTCTGTATCATGATGATAGCTATCTTCTAGACAAATTTCAACATCGG--GTCC 7179
QY 1730 gacaaaagaagaacccgtatccctaccctagctgtatgtcgtcggatt---agggcgg 1786
DB 7180 GCTGAAAGAGAAACCCATTAATCTTAACTGTAGCAGTCTTTTAGGATTTGGAGTGGCCG 7239
QY 1787 ccgtgtggtgaggaacagggacagctgcccctgatacagagcagcagcagctagagaaag 1846
DB 7240 CAGCTGGGGTAGTACTGGAACCGCTGCTTAATCCACACCCCGATCTTTGAG---- 7295
QY 1847 gactgtgtgagctacatcgcccatgacagagatctccgagccttaagaggtctgta 1906
DB 7296 -----GAATTAGCTACAGCTATGGTACTGTATCTCAGAGCTATAGAACACTCTATAA 7347

RESULT 12
US-08-716-351A-5
Sequence 5, Application US/08716351A
Patent No. 6033905
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based
TITLE OF INVENTION: Retroviral Vectors
NUMBER OF SEQUENCES: 5
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/716,351A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: PCT/US94/03784
APPLICATION NUMBER: PCT/US94/03784
FILING DATE: 06-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 15280-128-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10970 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..10970
OTHER INFORMATION: /standard_name= "p537 retroviral"
OTHER INFORMATION: vector
US-08-716-351A-5

Query Match 12.1%; Score 357.8; DB 3; Length 10970;

us-09-171-553b-9.feb22std.rni

Mon Feb 25 07:44:33 2002


```

1311 GAGAGTGCCTCACTCAACGATGAGACTGGAATGCGCCATCTCTCTC 1360
151 GlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTyrLys 167
1361 CAGGACCGGGTAAATTCCTCTTGTCAATTCGGCCCGGCGCACTACAA 1410
167 smethelysLeuTyrLysAspLysSerCysSerProSerAspLeuAspT 184
1411 AGTGATGAACATATATATAAGATAGAGCTGCTCCCATCAGACTAGATT 1460
184 TyrLeuLysLysSerPheThrGlyLysGlyLysGlnGluAsnLys 200
1461 ATCTAAGATAGATTCTCACTGAAAAAGAAAAACAGAAAAATATTCAAA 1510
201 TrpIleAsnGlyMetSerTyrGlyLysValPheTyrLysTyrGlyGly 217
1511 TGGATTAATGGTATGAGCTGGGGGAATGTTTTTAAATATGGGGGGG 1560
217 ValAlaGlySerThrLeuThrIleArgLeuArgLysGluThrGlyLys 234
1561 AGCAGAGTCCACTTAAACCAATTCGCTTAGATGAGAGGAGGAGACAG 1610
234 IProValAlaMetGlyProAspLysValLysAlaGlnGlyLysProPro 250
1611 CCCGTGTGGAGTGGGACCCGATAAAGTGGCTGGAACAGGAGGCGCCG 1660
251 AlaLeuGluProProHisAsnLeuProValProGlnLeuThrSerLeu 267
1661 GCCCTGGAGCCACCCATACCTTGGCGGTCGCCCAATTAACCTGCTCG 1710
267 GProAspIleThrGlnProProSerAsnSerThrThrGlyLeuLeuPro 284
1711 GCTTACATTAACACAGCGCTAGCAACGATGATGATGATGATGATGAT 1760
284 hrAsnThrProArgAsnSerProGlyValProValLysThrGlyLys 300
1761 CCAACAGCCCTAGAAACCTCCAGAGTGTCTCTTAAACAGACAGACAGA 1810
301 LeuPheSerLeuLysGlnGlyAlaPheGlnAlaIleAsnSerThrAsp 317
1811 CTCTTCACTTCATCCAGGAGCTTCCAGCCATCAACCTCCAGCCAGCC 1860
317 CAspAlaThrSerSerCysTrpLeuGlyLeuSerSerGlyProProTyr 334
1861 TGAATGCCACTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1910
334 TyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHis 350
1911 ATGAGGGGATGGCTAAAGAAAGAAATTCATGTGACCAAGAACATAGA 1960
351 AsnGlnCysThrTyrGlySerArgAsnLysLeuThrLeuThrGluVal 367
1961 AATCAATATGATGAGGGGTCCCGAAATAGCTTACCTCACTGAAATTC 2010
367 rGlyLysGlyThrCysLysLysLysAlaProProSerHisLysHisLeu 384
2011 CGGGAAGGGGACATGATAGGAAAGCTCCCATCCACCAACACCTTT 2060
384 yTyrSerThrValValTyrGlnGlnAlaSerGluAsnGlnTyrLeuVal 400
2061 GCTATAGACTGTGTTTATGACAGAGGCTAGAAAAATCAGATTATTA 2110
401 ProGlyTyrAsnArgTyrTrpAlaCysAsnThrGlyLeuThrProCys 417
2111 CCTGGTTATACAGGTGGTGGGATGCAATACCTGGTTAACCCCTGTGT 2160
417 LserThrSerValPheAsnGlnSerLysAspLeuGlyValMetValLys 434
2161 TTCACACCTCACTTCAACCAATCAAGAAATTCGTGTATGATGTCCAA 2210
434 LeuValProArgValTyrTyrHisProGlnGluValLysLeuAspGlu 450

```

```

2211 TCGTCCCCGAGTGTACTACATCTGAGGAAGTGTCTTGATGATAT 2260
451 AspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAl 467
2261 GACTATCGGTATTAACCCAGCAAAAGAAACCCGTATCCCTTACCTAG 2310
467 aValMetLeuGlyLeuGlyThrAlaValGlyValGlyTyrGlyThrAla 484
2311 TGTTATCTGTGGATTAGGGAGCGCGCTGGCTGATGAAACAGGAGAC 2360
484 ILeuThrThrGlyProGlnGlnLeuGlyLysGlyLeuGlyLysHis 500
2361 CCTGATACAGAGACACACAGCAGCTAGAGAAAGACTGTGATGATAC 2410
501 AlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAsn 517
2411 GCGCCATGACAGAAAGATCTCCAGGCTTGAAGAGAGTCTGTAGCA 2460
517 uGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArg 534
2461 AGAAGAGTCCCTGACTTCTTGTGTAAGTGTGTTCACAGAACCGAG 2510
534 LysLeuAspLeuLeuPheLeuArgLysGlyLysCysAlaAlaLeuLys 550
2511 GATTAGATCTGCTGCTTCTTAAGAAAGTGGTATGTCAGCCTTAA 2560
551 GluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSer 567
2561 GAAGAAATGTTGCTTATGATGATCACTCAGAGCCTCAGACATCA 2610
567 LAsnLysLeuArgLysLysLeuGluArgArgArgArgGluArgGluAla 584
2611 GACCAAGCTTAAAGAAAGTGTAGAGAGGCTCGAAGGAAAGAGAGCT 2660
584 sPcGlnGlyTrpPheGlnGlyTyrPheAsnArgSerProTrpMetThr 600
2661 ACCAGGGGTGTTGAAGATGTTTCAACAGGCTCTCTTGATGACAC 2710
601 LeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeu 617
2711 CTGCTTCTCTCTGACGGGACCCCTAGTAGTCTCTCTCTGTACTTAC 2760
617 rValGlyProCysLeuLysAsnArgPheValAlaPheValArgGluArg 634
2761 AGTTGGGCTTGTCTTAATTAATAGTGTGTGTGTGTGTGTGTGTGT 2810
634 aLserAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 650
2811 TGAATGACAGTCCAGATCATGTGTTAGGCAACAGTACCAAGCCTT 2860
651 SerGlnGlyGluThrAspLeu 657
2861 AGCCAAGGAGAAACTGACCTC 2881

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq.us-09-111-085-1
seq_documentation_block:
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2462

```

TYPE: DNA
ORGANISM: Porcine retrovirus
US-09-111-085-1

alignment_scores:
Quality: 2631.50 Length: 679
Ratio: 4.468 Gaps: 9
Percent Similarity: 86.745 Percent Identity: 73.049

alignment_block:

US-09-171-553B-10 x US-09-111-085-1 ..

Align seg 1/1 to: US-09-111-085-1 from: 1 to: 2462

```

1 MethSProthLeuSerThrParHISLeuProthArgGlyGlyLeuP 17
|||||
211 ATGATCCAGCTTAAGCGCGCCACCTCCGATTCGGGGTGGAAAGCC 260
17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaThrPheLeu 34
|||||
261 GAAAGACTGAAATCCCTTAACCTTCGCTCCATCGCGTGGTCTTA 310
34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSer 50
|||||
311 CTGTGTCATAGTACCTCAAGTAAAGTAAAGCCCTTGAGACAGCCG 360
51 AsnProHISArgProLeuSerProThrTrpLeuIleIleAspProAsp 67
|||||
361 AACGCCCTAAACCCCTTCTCACCCTGTTACTTACTGCTCCGCTAC 410
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
|||||
411 AGGATTAATATTAAACAGACTCAAGGAGAGCTCCCTTGAGGAGCTGGT 460
84 rProGlyLeuHISPheCysLeuArgLeuIleAsnProAlaVal..... 98
|||||
461 GGCCTGAATTAATATGCTGCTCGATCAGATCAATCCCGTCCATGAC 510
99 LysSerThrProProAsnLeuValArgSerTrpGlyPheTrpCysCysPr 115
|||||
511 CAGGCGACACCCCGATGATCTCGGTACCGGTTTAAACGTTTCC 560
115 oGly...ThrGlyLysGlyLysTrpCysGlySerGlyLysSerPheC 131
|||||
561 AGGACCCCAATATAGAGAAATATTGTGGAATCTCAGCATTTCTTTT 610
131 ysArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
|||||
611 GCAAGCAATGAGCTGCATACCTCTAATGATGGGAATGGAAATGGCCA 660
148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSer..... 161
|||||
661 GTCTCTCAGACAGACAGATGAGTACTCTTTTCTTAACAATCCTTACCAG 710
162 .....GLYProGlyLysTrpLys.....MetMetL 170
|||||
711 TTATATCATATTTAATTATGAGCATGGAGATGGAAGAGATTGGACAAGC 760
170 yLeuTrpLysAspLys.....SerCysSerProSerAsp 181
|||||
761 GGGTACAAAGATGTACGAATAAAGCAATTAACCTGTCATTCCTTAGAC 810
182 LeuAspTrpLeuLysIleSerPheThrGlyLysGlyLysGlnGluAsnI 198
|||||
811 CTAGATTACTTAAATAATAGTTTCACTGAAAAAGCAAAACAAGAAATAT 860
198 eGlnuSerTrpIleAsnGlyMetSerTrpGlyLysValPheTrpLysTrp 215
|||||
861 TCAAAAGTGGTAAATGATATCTTGAGGAATAGTAC.....TATG 904
215 LysGlyLysAla.....GlySerThrLeuThrIleArgLeuArgIle 228
|||||

```

```

905 GAGGCTGTGGAGAGAAAGATCTCTGTGACATATTCGCTCAGATA 954
229 GluThrGlyThrGluProProValAlaMetGlyProAspLysValLeu 245
|||||
955 GAACCTCAGATGGACCTCGGTGCTATAGACCAATATAGGTTTGGC 1004
245 agLingLysProProAlaLeuGluProProHISAsnLeuProValPro 262
|||||
1005 CGAACAAAGACCTCCATCCAAAGACAG..... 1032
262 InLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSerThr 278
|||||
1033 .....AGGCCATCTCTTAACCCCTCTGATTACATACCAACC 1068
279 ThrGlyLeuIleProThrAsnThrProArgAsnSerProGlyValPro 295
|||||
1069 TCTGGATCAATCCCACT.....GAGCCTAACATCTCACTAT 1103
295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAla 312
|||||
1104 TAAACAGGCGGCGAACTTTTAGCTCATTCAGGAGGCTTTTCAAGCTC 1153
312 LeAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSer 328
|||||
1154 TTAACTCCACGACTCCAGAGGCTACTCTTCTGTTGGCTTGCTTAGCT 1203
329 SerGlyProProTrpTrpGlyGlyMetAlaLysGlyLysGlyPheAsn 345
|||||
1204 TCGGGCCACCTTACTATAGAGGAATGGCTAGAGGAGGGAATCAATCT 1253
345 LThrLysGlnHISArgAsnGlnCysThrTrpGlySerArgAsnLysLeu 362
|||||
1254 GAACAAAGAACATAGAGAACCAATGATACATGGGATCCCAAAATACCTTA 1303
362 hrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProPro 378
|||||
1304 CCTTACTAGAGTCTCTGAAAGGCCACCTGCTAAGGATGGTCCCCCA 1353
379 SerHISGlnHISLeuCysTrpSerThrValValTrpGlyGlnAlaSerG 395
|||||
1354 TCCACCAACACCTTTGTAACACACACTGAAGCCTTAATACGAACCTCTGA 1403
395 uAsnGlnTrpLeuValProGlyLysAsnArgTrpTrpAlaCysAsnThrG 412
|||||
1404 GATTCATATACCTGTACCTGTTATGACAGTGIGTGCGCATTAACAG 1453
412 LLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspLeu 428
|||||
1454 GATTAAACCCCTGTGTTCCACCTTGCTTTCACCAAACTAAAGACTTT 1503
429 CysValMetValGlnIleValProArgValLysTrpHISProGlnGlu 445
|||||
1504 TGGGTTATGGTCCAAATGTGCCCCGGGTGACTACTATCCGAAAAAGC 1553
445 ValIleAspGlyTrpAspTrpArgTrpArgAsnArgProLysArgLysPro 462
|||||
1554 AGTCCTTGATGATATGACTATGATATATATGCGCAAAAGAGACCCA 1603
462 aISerLeuThrLeuAlaValMetLeuGlyLysGlyThrAlaValGlyVal 478
|||||
1604 TATCCCTGACACTAGCTGATATCTCGATTTGGAGATGGCTGAGGCGTG 1653
479 GlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGlyLysG 495
|||||
1654 GGAACAGGAGAGGCTGCCCTTAATCACAAGACCCCAACACTGGAGAAAG 1703
495 LysGlyGlyLeuHISAlaAlaMetThrGlyLysAspLeuArgAlaLeuLys 512
|||||
1704 ACTTAGTAACCTCATCATGATTTTAACGGAAGATCCCAAGCCCTGAAA 1753
512 InSerValSerAsnLeuGlnGlyLysSerLeuThrSerLeuSerGlyVal 528
|||||
1754 AATCTGTCACTAGCTGAGAGATCCCTTAACCTTATCTGAAGGTGT 1803

```

```

529 LeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgLuglyLyle 545
1804 CTACAGACAGAAAGGGGTTAGATCTGTCTTCTTAAAGAGAGAGGTT 1853
545 uCyAlaAlaLeuLysGluLysCysCysPheTyrValAspHisSerGlyA 562
1854 ATGTGTACCTTAAAGAGAGAAATGCTGCTCTATAGTACCTCAGAG 1903
562 LaIeArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArg 578
1904 CCATCAGAGACTCCATGACAGACTTAGAGAAAGGTTAGAGAGCGCTCA 1953
579 ArgLuarGluAlaAspGlnGlyTyrPheGluGlyTyrPheAsnArgse 595
1954 AGGGAAGAGAGAGGCTGACAGGGGTGCTTGAAGGATGGTTCAACAGGTC 2003
595 rProTyrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValVal 612
2004 TCCTTGATATACACCTGCTTCTGCTGACGGGCCCCCTAGTAGTCC 2053
612 euleuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 628
2054 TGCCTGCTTACTTACAGTGGGCTTGTATATATAGTTGTTGCC 2103
629 PheValArgLuarGlyValSerAlaValGlnIleMetValLeuArgLugl 645
2104 TTGTTAGAGACAGAGTGTGAGTGCAGTCCAGATCATGCTTAGCCACA 2153
645 nTyrGlnGlyLeuLeuSerGlnGlyLysIleuThrAspLeu 657
2154 GTACCAAGGCTTCTGAGCCCAAGAGAACTGACCTC 2190

seq_name: /cgn2_6/ptodata/2/ina/6B-COMB.seq:US-08-766-528-2

seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSER: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-2

alignment_scores:
    Quality: 2537.50      Length: 658
    Ratio: 4.452          Gaps: 9
    Percent Similarity: 86.626      Percent Identity: 72.796

alignment_block:
US-09-171-553b-10 x US-08-766-528-2 ..

Align seg 1/1 to: US-08-766-528-2 from: 1 to: 7333

22 IleProLeuSerPheAlaSerIleAlaTrpPheLeuThrIleTh 38
4806 ATCCCTTAAGCTTGCCCTCCATCCGCTGCTTCTTACTGTGCAATAC 4855
38 rProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHisArgp 55
4856 TCCTCAAGTAAATGTAACCGCTTGTGACAGCCCGCAACGCCATTAAC 4905
55 rLeuSerProThrTrpLeuIleIleAspProAspThrGlyValThrVal 71
4906 CCTTATCTCTCAGCTGTTACTTACTGACTCCGTTACAGGATTAATATT 4955
72 AsnSerThrArgGlyValAlaProArgGlyThrTrpTrpProGluLeu 88
4956 AACACACTCAAGGAGGAGCTCCCTGGGAGACCTGCTGCAATTA 5005
88 sPheCysLeuArgLeuIleAsnProAlaVal.....LysSerThrProp 103
5006 TGTCTGCTGATCATCAATCCCTGCTCAATGACAGCCACACCCC 5055
103 rOAsnLeuValArgSerTyrGlyPheTyrCysCysProGly...ThGlu 118
5056 CCGATGATCCGCTTACGCGGTTTACGTTTCCCGCAGGCCCAAT 5105
119 LysGluLysTyrCysGlyLysSerGlyLysSerPheCysArgArgTrpse 135
5106 AATGAGAATATTTGGGAATCCACAGATTTCTTTTCAAGCAATGAG 5155
135 rCyValThrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnA 152
5156 CTGCATTAATCTTAATGATGGAATGGAATGCGCAGTCTCAGCAAG 5205
152 sPArgValLysPheSerPheValAsnSer..... 161
5206 ACAGATAGTTACTCTTTGTAAACAATCTTACCAGTTATATCAATTT 5255
162 .....GlyProGlyLysTyrLys.....MetLysLeuTyrLysAs 174
5256 AATTATGSCCANTGAGAGATGGAAGATTTGGCAACAGCGGATCAAAAAGA 5305
174 pLys.....SerCysSerProSerAspLeuAspTyrLeu 186
5306 TGTACGAATAAGCAAAATATACCTTGTGATCTTATGATTAATCTTAA 5355
186 ySIIeSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrpIle 202
5356 AAATTAAGTTTCACTGAAAAAGAAACAGAAATATTCANAAATGGGTA 5405
203 AsnGlyMetSerTrpGlyIleValPheTyrLysTyrGlyGlyAla.. 218
5406 AATGATATCTTGGGGAATAGTGTAC.....TATGAGGCTCTGGGAG 5449
219 .....GlySerThrLeuThrIleArgLeuArgLuglyThrG 233
5450 AAAGAAAGATCTGTTCTGACTATTCCTCAGAAATGAGAACTCAGATGG 5499
233 LuProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyPro 249

```

5500 AACCTCGGTGCTATAGAGCAAAATAGGGTTGGCCGACAGACAGACT 5549
 250 ProAlaLeuGluProProHisAsnLeuProValProGlnLeuThrSerLe 266
 5550 CCATGCCAAGAACAG..... 5564
 266 uArgProAspIleThrGlnProProSerAsnSerThrGlyLeuIleP 283
 5565 .AGGCCATCTCCTAACCCCTGTGATTACATACAAACCTCGATGATGCTC 5613
 283 rofThsAsnThrProArgAsnSerProGlyValProValIysThrGln 299
 5614 CCACT.....GAGCCTAACATCATATTAAACAGGCGG 5648
 300 ArgLeuPheSerLeuIleGlnIleAlaPheGlnAlaIleAsnSerThas 316
 5649 AAACCTTTAGCCTCATCCAGGAGCTTTCAAGCTTTAACTCCACGAC 5698
 316 pProAspAlaThrSerSerCysTrpLeuGlyLeuSerGlyProT 333
 5699 TCCAGAGGCTACCTCTTCTTGGCTTGTAGCTTGGGCCACCT 5748
 333 YrTrGlnGlyMetAlaIysGluGlySphAsnValThrIysGluHis 349
 5749 ACTATGAGGGAATGCTAGAGAGGGAATTCATGTGACAAAGAACAT 5798
 350 ArgAsnGlnCysThrTrpGlySerArgAsnIysLeuThrLeuThrGluVa 366
 5799 AGAGAACATATACATGGGATCCCAAAATAGCTTACCTTCTAGAGGT 5848
 366 lSerIlysgIyThrCysIleGlyLysAlaProProSerHisGlnHisL 383
 5849 TTCTGAAAGGACACCTGATAGGATGTTCCCATCCACCAACACC 5898
 383 euCysTysSerThrValIyTrGlnIleAlaSerGluAsnGlnTyrLeu 399
 5899 TTTGTAACCACTGATCTTAAATCAACCTCTGAAGTCAATATCTG 5948
 400 ValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThrProC 416
 5949 GTACCTGGTTATGACAGGTGGGCGCATGTATATCTGATTAACCCCTTG 5998
 416 sValSerThrSerValPheAsnGlnSerLysAspLeuGlyValMetValG 433
 5999 TCTTCCACCTTGTTCACCAACAACTAAGACTTTGCTTATGTGTC 6048
 433 lIleValProArgValTyrTrHisProGlnGluValValLeuAspGlu 449
 6049 AAATGTCCCGGGGTACTACTATCCGAAAGAGCAGTCTGTATGAA 6098
 450 TyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLe 466
 6099 TATGACTATAGATATATCGGCCAAAAAGAGCCCATATCTCTGACACT 6148
 466 uAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThra 483
 6149 AGCTGTATGCTCGGATGGAGTGGCGTGCAGCGGTGGAGACAGAACGG 6198
 483 lAlaLeuIleThrGlyProGlnGlnLeuGlySglLeuGlyGluLeu 499
 6199 CTGCCCTTATCACAGCGGACAGCGGAGGAAAGAGACTTATGAACTTA 6248
 500 HisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAs 516
 6249 CATGCAATTTAGCGAAGATCTCCAGCCCTAAGAAAATGTGCAITAA 6298
 516 nLeuGlnIleuSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 533
 6299 CTTGAGGAATCCTTAACCTCTTATCTGAAGTGTCTTACAGAACAGAA 6348
 533 rGgIlyLeuAspLeuPheLeuArgGlnGlyGlyIleuGlySAlaAlaLeu 549
 6349 GGGGTTAGATCTGTATTCTTAAAGAGAGGAGGTTATGTGTAGCTTA 6398

550 LysGlnIuGysCysPheTyrValAspHisSerGlyAlaIleArgAspse 566
 6399 AAAGAGATGCTGTCTATGTATGATCATCTAGAGAGCCATCAAGACTC 6448
 566 lMetAsnLysLeuAlaGlyLysIleGluArgArgArgGluArgGluA 583
 6449 CATGAGCAAGCTTAGAGAAAGGTTAGAGAGCGTCGAAGGAAAGACAGG 6498
 583 lAspGlnGlyTrpPheGlnGlyTrpPheAsnArgSerProTyrMetThr 599
 6499 CTGACAGGGGTGGTTGAAGATGTTCAACAGCTCTCTTGGATGAC 6548
 600 ThrLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeu 616
 6549 ACCCTGCTTCTGCTGACGGGCCCCCTAGTACTCTGCTCTGTTACT 6598
 616 uThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluA 633
 6599 TACAGTTGGGCTTGTCTTATATAGTTTGTTCCTTGTATGAGAAC 6648
 633 rGValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 649
 6649 GAGTGAAGTGCAGTCCAGATCATGTACTTAGCGAACAGTACCAAGGCTT 6698
 650 LeuSerGlnGlyGluThrAspLeu 657
 6699 CTGAGCCAGAGGAAGTACACTGACTC 6722

seq_name: /cgn2_6/pdata/2/lna/6B_COMB.seq:US-08-766-528-1

seq_documentation block:
 : Sequence 1, Application US/08766528
 : Patent No. 6190861
 : GENERAL INFORMATION:
 : APPLICANT: Jay A. Fishman
 : TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
 : NUMBER OF SEQUENCES: 74
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: LAHIVE & COCKFIELD, LLP
 : STREET: 60 State Street
 : CITY: Boston
 : STATE: Massachusetts
 : COUNTRY: USA
 : ZIP: 02109-1875
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/766,528
 : FILING DATE:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/572,645
 : FILING DATE: 14-DEC-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Louis Myers
 : REGISTRATION NUMBER: 35,965
 : REFERENCE/DOCKET NUMBER: MGP-038CP
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (617)227-7400
 : TELEFAX: (617)227-5941
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 8060 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : US-08-766-528-1

alignment_scores:

Quality: 2402.00 Length: 655
 Ratio: 4.312 Gaps: 9
 Percent Similarity: 85.038 Percent Identity: 68.702

alignment_block:

US-09-171-553b-10 x US-08-766-528-1 ..

Align seg 1/1 to: US-08-766-528-1 from: 1 to: 8060

```

1 MethisProthrLeuSerTrpArgHisLeuProThrArgGlyGlyLeuPr 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
86 ATGCATCCACGCTTAAGCGCGCCACCTCCGATCGGGGTGGAAAGCC 135
17 OlysrArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeu 34
136 GAAAAGACTGAAATCCCTTAAGCTTCGCTCCATCGCGTGGTCCCTTA 185
34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 CTCGTCTCATATACCTCTCAGACTATAGTATGCGATGAGACAGCCTG 235
51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAspTh 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 AACTCCCATTAACCTTATCTCTCAGCTGTTAACTGACTCGCGCAC 285
67 rGlyAlaThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 AGGTATTTATATCAACAACACTCAAGGGAGGCTCTTAAAGAACCTGGT 335
84 rProGlnLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336 GCGCTGATCTATACGTTGCTCAGATCATGATATCTCTAGCTGACCTCA 385
101 ThrProProAsnLeuValArgSerTrpGlyPheTrpCysProGlyTh 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 ...CCCCAGATATCTCCATGCTCAGCGATTTATGTTGCCACGAGACC 432
117 rGlyLysGlu...LysTrpCysGlyGlySerGlyLysSerPheCysArg 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
433 ACCAATATATGAAACATGCGGAATCCAGACATTTCTTTGTAAAC 482
133 rGTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProLys 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
483 AATGGAACGTGTGTACTCTATATGATGATATGGAATGGCCAACTCT 532
150 LeuGlnAspArgValLysPheSerPheValAsn.....SerGlyProG 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
533 CAGCGAGATAGGGTAAGTTTCTTATGTCACACCTATACCAAGCTCTGG 582
164 LysTrpLysMetMetLysLeuTrpLysAspLysSer.....CysSerP 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 ACAATTTATATACCTGACCTGATGATGAACAGAACCCCAAGGCTCTC 632
179 rOserAspLeuAspTrpLeuLysIleSerPheThrGlyGlyLysGln 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
633 CTTTCAGACCTAGATTACCTAAATAATAGTTTCACATGAGAAAGAAAC 682
196 GlnAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTy 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683 GAAAATATCTTAATAATGGTAATGTAATGTTCTGGGAATGATATAT.. 730
212 rLysTrpGlyGlyGlyAla.....GlySerThrLeuThrIleArgL 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
731 ...TATGAGAGCTCGGGTAACACACAGGCTCCATCTTAACATATTCGC 776
226 euArgIleGlyThrGlyThrGlyProProValAlaMetGlyProAspLys 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
777 TCAAAATTAACACAG...CTGAGAGCTCCCAATGCTATAGACCAAAATACG 823
243 ValLeuAlaGlyGlnGlyProProAlaLeuGlnProProHisAsnLeuPr 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
824 GTCTGACGGGTCAAGACCCCAACCCCAAGACCA..... 859
259 ovalProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA 276
860 ..... 872
276 snSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
873 ACATTAACCTCTGATCAGACCCCACTGAGCTATGACAGCAGACT..... 916
293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaTh 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
917 .....AAATGGGGGCAAAATCTTTTATGCTTATCAGGAGCTTT 957
309 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeuc 326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
958 TCAAGCTCTTAACCTCCACGACTCCAGAGGCTTCTTCTGTGGCTAT 1007
326 LysLeuSerSerGlyProProTrpTrpGlyGlyMetAlaLysGlnArgLys 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1008 GCTTAGCTTGGGCCACACCTTACTATGAAGGAATGCTAGAGAGGAA 1057
343 PheAsnValThrLysGlnHisArgAsnGlnCysTrpTrpGlySerArgAs 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1058 TTCAATGTGACAAAGAACATAGAGACCAATGACATGGGGATCCAA 1107
359 nLysLeuThrLeuThrGlyValSerGlyLysGlyThrCysIleGlyLysA 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1108 TAAAGCTTACCTTACGAGGTTCTTGGAAGACACCTGCATGAGAAAG 1157
376 LysProSerHisGlnHisLeuGlySerThrValValTrpGlyGln 392
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1158 TTCCCCCATCCACCAACACCTTGTACCCACACTAACCCTTAATCA 1207
393 AlaSerGlnAsnGlnTrpLeuValProGlyTrpAsnArgTrpTrpLacy 409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1208 ACCCTGAAAGTCAATATCTGTAACCTGTTATGACAGGTGGGACAG 1257
409 sAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSerL 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1258 TAATACTGAGATTAACCCCTTGCTTCCACCTGTTTAAACCAACTA 1307
426 LysAspLeuCysValMetValGlnIleValProArgValTrpTrpHisPro 442
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1308 AAGATTTTTCATTAATGCTCAATGTTCCCAAGTGTATTAATCAATCC 1357
443 GlnGlyValAlaLeuAspGlnTrpAspTrpArgTrpArgTrpLysAsr 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1358 GAAAACCAATCTCTGATGATGATGACTACAGAAATCATGACAAAGAG 1407
459 rGlyProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaV 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1408 AGAACCCATATCTCTGACACTTCTGATGATGCTGAGACCTGGAGTGCAG 1457
476 aGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1458 CAGGTGAGAGACAGAACAGCTGCTGCTGACAGGACACAGCAGCTA 1507
493 GlnLysGlyLeuGlyGlnLeuHisAlaAlaMetThrGlnAspLeuArgL 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1508 GAAACAGACTTGTATACCTACATCCGATTTGTAACAGAGATGCCAAGC 1557
509 aLeuLysGlnSerValSerAsnLeuGlnGlnSerLeuThrSerLeuSerG 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1558 CTTAGAAATAATCTGTAGTAACCTGGAGGAATCCCTTAACCTCTTATCTG 1607
526 LysValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgL 542
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1608 AAGTAGCTCTACAGAAATGAAGAGGTTAGATTATATATTTCTAAAGAA 1657
543 GlyLysLeuCysAlaIleLeuLysGlnLysCysCysPheTrpValAspH 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1658 GAGAGATTATGTGTAGCTTGAAGAGAGAAATGCTGTTTATATGTGATCA 1707

```

```

559 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluA 576
|||||
1708 TTCAGGGGCCATCAGAGCTCCATGACAACTAGAGAAAGTTGGAGA 1757
|||||
576 rgaArgArgArgGluArgGluAlaAspGlnGlyTrpPheGlyTrpPhe 592
|||||
1758 AGCGTCAGAGGAAAGAACTACTCAAGGCTGGTGGAGGATGCTTC 1807
|||||
593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyPro 609
|||||
1808 AACAGGCTCCTTGTTGGCTACCTTACTTCTGCTTTAAACAGACCTT 1857
|||||
609 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuLeuAsnArg 626
|||||
1858 AATAGTCCTCCTCCTTACTCAAGTGGCCATGATATTATTAAACAAT 1907
|||||
626 heValAlaPheValArgGluArgValSerAlaValAlaIleMetValLeu 642
|||||
1908 TAAATGCTTCATTAGAGAGCAATAGTGCAGCTCCAGATCATGTAATT 1957
|||||
643 ArgGlnGlnTrpGln 647
|||||
1958 AGAACACAGTACCA 1972
|||||

seq_name: /cgn2_6/ptdata/2/ina/6B_COMB.seq:US-08-766-528-3

seq_documentation_block:
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: WGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-3

alignment_scores:
Quality: 2395.00      Length: 655
Ratio: 4.292          Gaps: 10

```

```

Percent Similarity: 85.191      Percent Identity: 68.855
alignment_block:
US-09-171-553B-10 x US-08-766-528-3
Align seq 1/1 to: US-08-766-528-3 from: 1 to: 8132

1 MethisProHisLeuSerTrpArgHisLeuProThrArgGlyGlyIleuPr 17
|||||
5620 ATGCATCCACAGTTAAACCGCGCCACCTCCGATTCGGGGTGAAGACG 5669
|||||
17 OlyArgLeuArgGlyLeuLeuSerPheAlaSerIleAlaTrpPheLeu 34
|||||
5670 CAAGAAGCTAATAATCCCTTAAGCTTGCGCTCCATCGCTGGTTCCTTA 5719
|||||
34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspProAsp 50
|||||
5720 CTCTGTCAATTAACCTCTCAGACTAATGTAATGGCATAGACAGACCTG 5769
|||||
51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAsp 67
|||||
5770 AACTCCCATAAACCTTATCTCTCACCTGTGTAATTACTGACTCCGCGAC 5819
|||||
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
|||||
5820 AGGTATTATATATACACACACTCCAGGAGGAGCTCTTTAGAGAACCTGT 5869
|||||
84 rPrProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
|||||
5870 GGCCTGATCATACGTTTGCCTCAGATCAAGTATTCATGCTGACCTCA 5919
|||||
101 ThrProProAsnLeuValArgSerTrpGlyPheTrpCysProGlyTh 117
|||||
5920 ...CCCCAGATATCCCTCATCTCCAGCATTTTATGTTGGCCAGGAC 5966
|||||
117 rGluLysGlu...LysTrpCysGlyGlySerGlyLeuSerPheCysArg 133
|||||
5967 ACCAATTAATGGAACAAATGGCGAATCCCGAGACTTTCTTTGTAAAC 6016
|||||
133 rGTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
|||||
6017 AATGGAACCTGTAAACCTCTTAATGATGATATTGAATGGCCAACTCT 6066
|||||
150 LeuGlnAspArgValLysPheSerPheValAsn.....SerGlyProG 164
|||||
6067 CACGACGATAGGGTAAGTTTCTTATGTCAACACACTATACCACTCTGG 6116
|||||
164 LysTrpLysMetLeuLysLeuTrpLysAspLysSer.....CysSerP 179
|||||
6117 ACAATTATTAATTACCTGACCTGATTAAGACAGGAAAGCCCAAGTCTCTC 6166
|||||
179 rOserAspLeuAspTrpLeuLysLysSerPheThrGluLysGlyLysGln 195
|||||
6167 CTTCAAGACCTAGATTACCTAAATAAATTAAGTTTCACCTGAGAAAGCAACAA 6216
|||||
196 GluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPhe 212
|||||
6217 GAAATATATCCTTAATAATGGTAAATGATGCTGGGAAAGGATATAT.. 6264
|||||
212 rLysTrpArgGlyGlyAla.....GlySerThrLeuThrIleArgL 226
|||||
6265 ...TATGAGAGCTCGGCTAAACAACAGGCTCATCTTAACATTATGCC 6310
|||||
226 euArgIleGlnThrGlyThrGluProProValAlaMetGlyProAspLys 242
|||||
6311 TCATAATTAACCAAG...CTGAGAGCTCAATGGCTATAGACCAATAATACG 6357
|||||
243 ValLeuAlaGlnGlnGlyProProAlaLeuGlnProProHisAsnLeuPr 259
|||||
6358 GTCTTGACGGGTCAAAAGACCCCAACCAAGACCA..... 6393
|||||
259 oValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA 276
|||||

```


576	rgatgataatgguatgctunlaaspnglytrpphegluglytrphe	592
7292	agccgcgaaaggaaaaagaaactactcaagggtgctttgacgcatgcttc	7344
593	asnatsarprotpmetthrthtleuensealaleuthrglyprole	609
7342	aacaggtctcttggsttgacacacctactcttctgtttaaacagaccctt	7391
609	valvalleuleuleuleuleuthrvalglyprotyseulleaasnacrp	626
7392	aatatctccctccctcgatctactcacagttgggcacatgtrttatatacagst	7441
626	hevalalaphhevalatrgsluargvalsealavalglnilemetvalleu	642
7442	taattgacctctattagagacaaataaagtcagctcaatcatggtacttt	7491
643	atrgslngintyrgln	647
7492	agacacacagttaccaa	7506

7492 AGACAACAGTACCAA 7506

[illegible]

Align seg 1/1 to: US-09-376-781-25 from: 1 to: 2000

[illegible]

193 AACGCCATTAACCCCTTATCTCTGACCGCTGTAATTACTAGCTCTGCAC 242
67 rglyValThValAsnSerThArglyValAlaProArglyThTrpT 84
243 AGGATTAATATCCACAGCGCTCGAGGGAGCGCTCTTAAAGAAACCGGT 292
84 rProGluHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
293 GGCGTATCTATATGTCTGCTCCCTCAATGATGATTCCTAGCTGCACCTCA 342
101 ThProProAsnLeuValArgSerIglyPheThyCysCysProGlyTh 117
343 ACC...CCAGATATCTCCGCTGCTTACGAGATTTATGTTGGCCAGGACC 389
117 rGluLysGlu...LysTyrCysGlyIglySerGlyGluSerPheCysArg 133
390 ACCAAATAAATGAAAAACATATAGAAACCTAGAGATTTCTTTTACAAAC 439
133 rgrTsrSerCysValThrSerAsnAspGlyAspTrpLysTrpPoliSer 149
440 AAGGAGCGCTGTAACTCTATATGATGAAATGCAAAATGGCCACCTCT 489
150 LeuGlnAspArgValLysPheSerPheValAsnSerIglyProGlyLysTy 166
490 CTCAGAGATAAGGTAAAGCTTTCTTATCTATCCAAC...CCCAATACCA 533
166 rLysMetLeuLysLeu...TyrLysAspLysSer...CysSerProSerA 181
534 CCGGACCTGGAAACGCTCATACAGGTACCAATTCGGGTGTTTCCCTCAG 583
181 sPLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsn 197
584 ACCTAATATATCTTAAATTAAGTTTCAACGAAAAAAAACAAGAAAT 633
198 IleGlyLysTrpIleAsnGlyIleMetSerProGlyIleValPheThyLys 213
634 ATCTTAATAATGCATTAATAGTATGTCTCGGGGAATAATATATATACAG 683
214 TyrGlyIglyValGlyIleArgLysSerThrLeuThrIleArgLeuArgIleGluT 230
684 TTGGGACAGACAAACAGGCTCATTTACCATTCACCATCCGACTTAATAA...A 730
230 hrcGlyhrcLupProProValAlaMetGlyProAspLysValLeuAlaGlu 246
731 GCACGCTAAAGCCTCAATGGCTATAGAGACGAATACGGCTTAAAGCGGT 780
247 GlnGlyProProAlaLeuGluLupProProHisAsnLeuProValProGlu 263
781 CAAGAAGACCCCAACCCACAGACCATCTCTCATATA... 816
263 uhrSerLeuArgProAspIlePhrGlnProProSerAsnSerThrThg 280
817 ...ACTTCTAATTAAGACCCACACTGAG...TCTAACAGCAGACT. 855
280 lyeuLierProHisnThrProArgAsnSerProGlyValProValLys 296
856AAA 858
297 ThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAs 313
859 ACGGGAGCAAAACCTTTTAGTCAATCAGAGAGACTTTTCAACGCTCTAA 908
313 nSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerG 330
909 CTCACAGACTCCAGAGGCTACCTCTTCTGTGTGGCTTTGCTTAACTTCG 958
330 lYProProTyrTyIglyLysMetAlaLysGluArgLysPheAsnValThr 346
959 GCCACACTTACTATATAAGAAATGGCTAAAGAGAAAAATTCAATGTGACA 1006
347 lYsGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLe 363

1009	AAAAACATGAGAACCAATGATCGGGATGCCAAATTAAGCTTAACCT	10358
363	urhrgluValSerGlyLysGlyIYthrcysIleGlyLysAlaProSerH	380
1059	TACTGAGCTTCTTGGAAGACACCTGCATATAAAAAAGGTTCCCATGCC	11088
380	isclnHisLeuCystrSerThrValValTYGluGlnAlaSerGluasn	396
1109	ACCAACACCTTTAACCAACACACTGAAGCCTTTAACTCAAAACCTGAGAGT	11588
397	GlnTYrLeuValProGlyTyrasnArgprTPalacyAsnThrGlyLe	413
1159	CAATATCTGCTACTCTGTTATGACAGGtGGtGGCATGTAATACGGGAT	12088
413	urhProCysValSerThrSerValPheasnInSerIysAspLeuCyV	430
1209	AAACCTTGtGTTCACACTGCTGGTTTTCACCAACAACTAAAGCTTTTACA	12588
430	alMeValGlnIleValProArgValTYrTYrHisProGluGluValAl	446
1259	TYrATGGCCAAATTGTTCCCCAGATATATTACTATCCCAAGAAACAAAT	13088
447	LeuAspGlyTYrAspTYrArgTYrArgTYrAsnArgProLysArgGluProVal	463
1309	CTGATGATATATGATTAACAGGAACCATCATCAAAAAAGAAACCCATATTC	13588
463	rIeuThrLeuAlaValMetLeuGlyIleuGlyTYrThrAlaValGlyAl	480
1359	CTGACACCTGCGAGTATGCTGGACCTCGAGATGATTAACAGTGTGAGAA	1408
480	hrcIlyThrAlaAlaLeuIleThrgIlyProGlnGluGluGlyLeu	496
1409	CAGGAACCTCAGCTTATGATTACAGGACCTCAGACACTGAAACAGGACTT	14588
497	GlyGluLeuHisAlaAlaMetThrcGluaspLeuArgAlaLeuGlyGlu	513
1459	AGTAACCTTCATCAATAATGTATACAGGAATCTCCAAAGCCCTTAATAAAATC	15088
513	rValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuG	530
1509	TGTCACTAATCCTGGAATAATCCCTAACCCCTTACTGTAATAGTGTAC	15588
530	hAsnArgArgIleuAspLeuaspLeuLeuValGlyGlyLeuGlyCys	546
1559	AGATATAAAAAGGGTTAGATTATATTATTCTAAAAAAGAAGATTATGT	16088
547	AlaAlaLeuLysGluGluCysGlySerIleYrValAspHisSerGlyAlaI	563
1609	GTAACCTTTAAAGAGAAATAGCGTTTATATATGATCATCTTAAGGGGCAT	16588
563	eArgAspSerMetAsnLysLeuArgLysIleuGluLuarArgTrArgArg	580
1659	CAGAGCTCCTCATGACCAACCTTAAAAAAGGTTGAGAGAACCTGGAAGG	17088
580	hArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerPr	596
1709	AAAAGAAACTTACTCAAAAGATGTTTAAAGATGGTTCAACAGGCTGCC	17588
596	orPrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuL	613
1759	TTTGTTGGGTCCCTACTTCTTACTTTTAACAGGACCCCTTAATGCTGCC	18088
613	euleuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPhe	629
1809	TCTCTTACTCATCAGTTGGCCATGTATTATTAAACAGTTATATGCTTCC	18588
630	ValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGlnTY	646
1859	ATTAGAAACGAAATTAAGCAGCTCCAGATCATGTGACTTGAACAACAGTA	19088
646	rGln 647	
1909	CCAA 1912	

seq_name: /cgn2_6/plodata/2/ina_6B_COMB.seq-us-09-376-781-30

seq_documentation_block:

; Sequence 30, Application US/09376781

; Patent No. 6261806

; GENERAL INFORMATION:

; APPLICANT: Banerjee, Papia T.

; APPLICANT: Patience, Clive

; APPLICANT: Andersson, Goran K.

; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of

; Patent No. 6261806

; TITLE OF INVENTION: Use

; FILE REFERENCE: 61750-267

; CURRENT APPLICATION NUMBER: US/09/376,781

; CURRENT FILING DATE: 1999-08-18

; EARLIER APPLICATION NUMBER: 60/097,015

; EARLIER FILING DATE: 1998-08-18

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 30

; LENGTH: 2000

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Nucleotide

; OTHER INFORMATION: consensus sequence of PERV-D.

US-09-376-781-30

alignment_scores:

Quality: 2169.00

Ratio: 4.032

Percent Similarity: 82.515

Percent Identity: 64.571

alignment_block:

US-09-171-553b-10 x US-09-376-781-30 ..

Align seg 1/1 to: US-09-376-781-30 from: 1 to: 2000

```

1 MethisProThleuSerTrpArgHisLeuProThrArgGlyGluIrr 17
  ::::::::::::::::::::
43 ATACATCCACGCTTAAGCCGCGCCACCCCTGATTCAGGGTGAAGGCC 92
17 olyAsArgLeuArgLeuProLeuSerPheAlaSerIleAlaTrpPheLeu 34
  ::::::::::::::::::::
93 AAAAAGACTAAAATCCCTTAAGCTTCGCTCCATCAGATGATGCTCTTA 142
34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
  ::::::::::::::::::::
143 CTCTGTCAATTAACCTCTCAGACTAATGATGATGATGATGATGATGATG 192
51 AsnProHisArgProLeuSerProThrTrpPheIleIleAspProAspH 67
  ::::::::::::::::::::
193 AACTCCCAATAAACCTTATCTGACCTGTTAATTAATGATGATGATGATG 242
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
  ::::::::::::::::::::
243 AGGTATTAATATCCACAGCGCTCGAGGAGGAGGCTCTTAAAGACTGCT 292
84 rPProGlnLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
  ::::::::::::::::::::
293 GGCTGATCATATGATGCTGCTCAGATGATGATGATGATGATGATGATG 342
101 ThrProProAsnLeuValArgSerTrpGlyPheTrpCysCysProGly 117
  ::::::::::::::::::::
343 ACC...CCAGATATCCCTCGTCTAGGATTTATGTTGGCCAGGACC 389
117 rGlyLysGlu...LysTrpCysGlyGlySerGlyGlySerPheCysArg 133
  ::::::::::::::::::::
390 ACCAATTAATGAAGAACTATGAAATCTAGAGATTTCTTTTACAAAC 439
133 rGTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149

```

```

440 AATGAGACTGTGTAACTTAATGATGAAATGAAATGCAACCTCT 489
150 LeuGlnAspArgValLysPheSerPheValAsnSerLysProGlyLys 166
  ::::::::::::::::::::
490 CTGCAAGATAGGTAAGCTTTCTTATGTCAAC...CCCAATATACCA 533
166 rLysMetLeuLysLeu...TyrLysAspLysSer...CysSerProSerA 181
  ::::::::::::::::::::
534 CCGGACCTGGAAACGTCATACAGAGTACCAATTCGGGTGTTTCCCTCAG 583
181 sPLeuAspTrpLeuLysIleSerPheThrGluLysGlyLysGlnGluAsn 197
  ::::::::::::::::::::
584 ACTAGATTTATCTTAAATAGTTTACCGAAAAAACAAGAAAT 633
198 IleGlnLysTrpIleAsnGlyMetSerTrpGlyLysLeuAlaPheTrp 213
  ::::::::::::::::::::
634 ATCTTAAATGATTAATGATGATGCTGCGGAAATATATATATATACAG 683
214 .TyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGlu 230
  ::::::::::::::::::::
684 TTGGGACAGACAAACCGCTCCATTTAACCATCCGACTTAATAA...A 730
230 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 246
  ::::::::::::::::::::
731 GCCAGTAAAGACCTCAATGAGCTTATGAGACCAATATAGCTTTAACGG 780
247 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGln 263
  ::::::::::::::::::::
781 CAAGAACCCTCAACCCAGGACCTCTGATATA... 816
263 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrArg 280
  ::::::::::::::::::::
817 ...ACTTCTAATTAAGCCCACTAG...TTAACAGCAGCAGCT. 855
280 LysLeuIleProIleAsnThrProArgAsnSerProGlyValProVal 296
  ::::::::::::::::::::
856 .....AAA 858
297 ThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 313
  ::::::::::::::::::::
859 ACGGGGACAAACTTTTATGCTATCTCAGAGACTTTTACAGCTCTTAA 908
313 nSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSer 330
  ::::::::::::::::::::
909 CTCACAGACTCCAGAGGCTACCTCTCTGTTGCTTCTTACCTACGCG 958
330 LysProProTrpTrpGlyGlyMetAlaLysGluArgLysPheAsnValThr 346
  ::::::::::::::::::::
959 GCCACACTTACTATTAAGAAATGGCTAAAGAGAAATTCATGATGACA 1008
347 LysGlnHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThr 363
  ::::::::::::::::::::
1009 AAAAAGACTTAAGACCAATGATGATGAGGATCCCAAAATTAAGCTTAC 1058
363 uThrGlnValSerGlyLysGlyThrCysIleGlyLysAlaProProSerH 380
  ::::::::::::::::::::
1059 TACGAGGTTCTCGAAAGACACCTGATTAATAAAGGATCCCATCC 1108
380 LysGlnHisLeuCysTrpSerThrValValLysGlnGlnAlaSerGluAsn 396
  ::::::::::::::::::::
1109 ACCAAGACCTTTTACACACACCTGATTAATTAACCTGAGAGT 1158
397 GlnTrpLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnTrpGly 413
  ::::::::::::::::::::
1159 CAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1208
413 uThrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCys 430
  ::::::::::::::::::::
1209 AACCCCTGTGTTCACACTGTTTCAACCAAACTAAAGACTTTTACA 1258
430 alMetValGlnIleValProArgValLysTrpHisProGlnGluValVal 446
  ::::::::::::::::::::

```

```

1259 TTAAGTCCAAATGTTCCCGAGATATTACTATTCACCAAGAAACAAATT 1308
447 LeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSe 463
|||||
1309 CTGATGATATATGATTACAGGAACATCCGCAAAAGAAAACCCCATATC 1358
463 rLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyT 480
|||||
1359 CCGACACTCGCAGTAATGCTCGACTCGAGATACAGAGTGTAGAGA 1408
480 hrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluysGlyLeu 496
|||||
1409 CAGGAACCTGCGCTTGTAGTACAGGACCCGACGAGCTAGAACAGACATT 1458
497 GlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuGlySe 513
|||||
1459 AGTAACCTACATCAATTTGTACAGAAATTCACAGCCCTTAAAAAAATC 1508
513 rValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuG 530
|||||
1509 TGTCACTAACCTGGAATAATCCCTACCTTATCTGAAGTATCTTAC 1558
530 LAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyLeuGlyLeu 546
|||||
1559 AGATATAAAAAGGTTAGATTATTTCTAAAAAAGAAAGATATTATGT 1608
547 AlaAlaLeuLysGluGluLysCysPheTyrValAspHisSerGlyAlaAl 563
|||||
1609 GTAGCCTTAAAGAGAAATGCTGTTTATGTAGATCATTCAGGGGCCAT 1658
563 eArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArgArg 580
|||||
1659 CAGAGACTCCATGAAACACCTTAAAAAAGTTGGAGAAACGTGCAAGGG 1708
580 LArgGluAlaAlaAspGlnGlyTyrPheGluGlyTyrPheAsnArgSerP 596
|||||
1709 AAAAGGAAACTTACTCAAGATGGTTTAAAGATGTTTAAACAGGCTTCC 1758
596 oTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 613
|||||
1759 TTGGTTGGCTACCTTACTTCTTAAACAGACCCCTTAAATAGTCTTCC 1808
613 eLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPhe 629
|||||
1809 TCTGTACTACTCAGAGTTGGGCATGATATTAACAAGTTAATGCTTCC 1858
630 ValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGly 646
|||||
1859 ATTAGAAACGAATAAGTGCAGTCCAGATCATGTAAGTACAAACAGTA 1908
646 rGln 647
|||||
1909 CCAA 1912

```

seq_name: /cgn2_6/prodata/2/ina/6B-COMB.seq:us-09-376-781-24

seq_documentation_block:

Sequence 24, Application US/09376781

Patent No. 6261806

GENERAL INFORMATION:

APPLICANT: Banerjee, Papia T.

APPLICANT: Patience, Clive

APPLICANT: Andersson, Goran K.

TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of

Patent No. 6261806

TITLE OF INVENTION: Use

FILE REFERENCE: 61750-267

CURRENT APPLICATION NUMBER: US/09/376,781

CURRENT FILING DATE: 1999-08-18

EARLIER APPLICATION NUMBER: 60/097,015

EARLIER FILING DATE: 1998-08-18

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 24
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: sequence of the 3' end of the PERV-D env region.
US-09-376-781-24

alignment_scores:
  Quality: 1675.00      Length: 489
  Ratio: 4.095          Gaps: 8
  Percent Similarity: 83.640  Percent Identity: 66.871

alignment_block:
US-09-171-553B-10 x US-09-376-781-24 ..

Align seg 1/1 to: US-09-376-781-24 from: 1 to: 1493

163 ProGlyLysTyrLysMetMetLysLeu...TyrLysAspLysSer...Cy 177
|||||
4 CCAATTAACACCGGACCTGGAAAACTCATACAGGTACCATTTCTGGGATG 53
177 sSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluTyrGlyL 194
|||||
54 TTTTCCCTCAGACCTAGATTATCTTAAATATAGTTTCACGCAAAAAAAA 103
|||||
194 ysgIngluAsnIleGlnLysTyrPilleasnGlyMetSerTyrPglyIleVal 210
|||||
104 AACAAAGAAATATCTTAAATGATGATAATGATGATGCTCGGGGAATATA 153
|||||
211 PheTyrLys...TyrGlyGlyAlaGlySerThrIleuThrIleArgLe 226
|||||
154 TATTATACAGGTGGGACAGCAACACGAGCTCCATTCTAACACATCCGACT 203
|||||
226 uArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLysV 243
|||||
204 TAAATA...AGCCAGCTAGAGCCTCCATGCTATAGAGACGGAATACGG 250
|||||
243 alLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPro 259
|||||
251 TCTTAACGGGTCAAGAAACCCCAACCCAGGACCATCCTCGATATA... 297
|||||
260 ValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerAs 276
|||||
298 .....ACTTCTAAATTAAGACCCACAGTGA.....TCTAA 326
|||||
276 nSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGlyV 293
|||||
327 CAGCAGCAGCT..... 336
|||||
293 alProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPhe 309
|||||
337 .....AAAACGGGACCAAAACTTTTGTAGTCAATCCAGGAGGCTTT 378
|||||
310 GlnAlaIleAsnSerThrAspProAspAlaThrSerSerGlyTyrPleuCy 326
|||||
379 CAAGCTCTTAACCTCAGACCTCCAGAGGCTACCTCTTCTGTGGCTTTG 428
|||||
326 sLeuSerSerGlyProProTyrTyrGluGlyMetAlaValGluArgLysP 343
|||||
429 CTTAATCTGGGGCCACCTTACTTATAGAAAATGCTTAAAAAGAAAAAT 478
|||||
343 heAsnValThrLysGluHisArgAsnGlnCysThrTyrGlySerArgAsn 359
|||||
479 TCAATGTGACAAAAAACAATAGACCAATGATCATGCGGATCCCAAAAT 528
|||||
360 LysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAl 376
|||||
529 AAGCTTACCTTACTGAGGTTTTCGAAAGAACCTCTTAAAAAAGGT 578
|||||
376 aProProSerHisGlnHisLeuGlySerTyrSerThrValValGluGlnA 393

```

```

579 TCCCCATCCCAACCAACCTTTACCAACACACTGAAGCCTTTATATCAA 628
339 lSserGluasnGlnTyLeuValProGlyTyraAsnArgTPTrPalaCys 409
629 CCTGTGAGAGTCANATATCTGGTACCTGGTATATACAGAGTGGTGGCATGT 678
410 AsnThyGlyLeuThrProCysValSerThSerValPheAsnGlnSerIly 426
679 AATACAGATTAACCCCTGTGTGTTCACCTTGCGTTGATTCACCAACATAA 728
426 sasPleuCysValMetValGlnIleValProArgValTyTrPheIsProg 443
729 AGACTTTTACATATATGATGTCCAAAATGTGTCCCGAGTATATTACTATCCA 778
443 LucIluValValLeuAspGluTyraAspTyraArgTyraAsnArgProLyArg 459
779 AGAAACCAATTCCTCGATGATATGATTACAGGAACCATGCACAAACAAAA 828
460 GluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaVal 476
829 AAACCATATCCCTGCACACTCGCAGATATCTCGACTCGCAGAGATAC 878
476 lglValAlGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuG 493
879 AGGTGTGACAAACAGGAACCTGCACCTTATGTACAGGCACCTCAGCAGTAG 928
493 luluGlyLeuGlnGlyIleuHisAlaAlaMetThrGluAspLeuArgAla 509
929 AAACAGGACTTACTAACCTTACATCAATTTGTACAGGAATTTCCAAACC 978
510 LeuIlyGluSerValSerAsnLeuGlnGluGlySerLeuThrSerLeuSerG 526
979 CTAAAAAAATCTGTCTAGTAACTCGAATAAAATCCCTTACCTCTATCTGA 1028
526 uValValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGln 543
1029 AGAGTTCTTACACAAATAAAAAGGGTTAGATTATTATTTCTAAAAAAA 1078
543 lYcIlyLeuCysAlaAlaLeuIlyGlnGluCysCysPheTyValAspHis 559
1079 GAAGATTATGTGTAGACCTTAAAGAGAAATGCTGTTTPTGTGTGATCAT 1128
560 SerGlyAlaIleLeuArgPseSerMetAsnTyLeuArgTyraGlyLeuArg 576
1129 TCAGGGGCCATCAGAGCTCTCATGAAACAACTTAAAAAAGGTGGAGAA 1178
576 gATgATgATgATgATgATgATgATgATgATgATgATgATgATgATgATg 592
1179 ACCTGCAAGGAAAGAAAGAAACCTTACCTCAAAAGATGCTTTAAAGATGCTC 1228
593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLe 609
1229 AACAGGTCTCTCTGTGGTTCACCTTACTTCTTAAACAGGACCCCT 1278
609 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgP 626
1279 AATAGTCTCTCTCTGTGTACTCAGAGTGGGCCATGATATTATTAACAAGT 1328
626 heValAlaPheValArgGluArgValSerAlaValGlnIleMetValLeu 642
1329 TATATCTCTTCAATTAGAAACAGATATAGTCCAGTCCAGATCATGTGCTACT 1378
643 ArgGlnGlnTyTrIln 647
1379 AGCAACACAGTACCA 1393
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-011-745-7
seq_documentation_block:
; Sequence 7, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:

```

```

1  APPLICANT: Collins, Mary KL
2  APPLICANT: Weiss, Robin A
3  APPLICANT: Takeuchi, Yasuhiro
4  APPLICANT: Cosset, Francois-Loic
5  TITLE OF INVENTION: Expression systems
6  FILE REFERENCE: 09/011,745
7  CURRENT APPLICATION NUMBER: US/09/011,745
8  CURRENT FILING DATE: 1998-06-22
9  EARLIER APPLICATION NUMBER: PCT/GB96/02061
10 EARLIER FILING DATE: 1996-08-23
11 EARLIER APPLICATION NUMBER: GB9517263.1
12 EARLIER FILING DATE: 1995-08-23
13 NUMBER OF SEQ ID NOS: 29
14 SOFTWARE: PatentIn Ver. 2.0
15 SEQ ID NO 7
16 LENGTH: 6312
17 TYPE: DNA
18 ORGANISM: Artificial Sequence
19 FEATURE:
20 OTHER INFORMATION: Description of Artificial Sequence: Portion of
21 FEATURE:
22 OTHER INFORMATION: construct
23 NAME/KEY: misc_feature
24 LOCATION: (4058)
25 OTHER INFORMATION: n is any nucleotide
26 FEATURE:
27 NAME/KEY: misc_feature
28 LOCATION: (4059)
29 OTHER INFORMATION: n is any nucleotide
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: (4060)
33 OTHER INFORMATION: n is any nucleotide
34 FEATURE:
35 NAME/KEY: misc_feature
36 LOCATION: (4061)
37 OTHER INFORMATION: n is any nucleotide
38 FEATURE:
39 NAME/KEY: misc_feature
40 LOCATION: (4246)
41 OTHER INFORMATION: n is any nucleotide
42 FEATURE:
43 NAME/KEY: misc_feature
44 LOCATION: (4247)
45 OTHER INFORMATION: n is any nucleotide
46 FEATURE:
47 NAME/KEY: misc_feature
48 LOCATION: (4248)
49 OTHER INFORMATION: n is any nucleotide
50 FEATURE:
51 NAME/KEY: misc_feature
52 LOCATION: (4249)
53 OTHER INFORMATION: n is any nucleotide
54 US-09-011-745-7

```

```

alignment_scores:
    Quality: 1544.50      Length: 660
    Ratio: 3.300         Gaps: 15
    Percent Similarity: 70.909      Percent Identity: 46.970

alignment_block:
    US-09-171-553B-10 x US-09-011-745-7 ..

Align seg 1/1 to: US-09-011-745-7 from: 1 to: 6312

47 ILeaSPSeRseRanSProHISaRqProlenSeSProHrThrPleuIleI 63
   :::::::::::|||||:::|||||:::|||||:::|||||:::
1675 CTGCAAAATAGAAACCCCGACGAGCCATGACCTCCTACTGGCAGTACT 1724
   :::::::::::|||||:::|||||:::|||||:::|||||:::
63 eaSPProaSPThncIyAlthrValaInSeSrThrARgIyAlaIaPro 80
   :::::::::::|||||:::|||||:::|||||:::|||||:::
1725 GTGCC...CAAACTGGAGACCTGTCTGGATACCAAGGACGTGAGCCCC 1771

```

```

80  rglYThrTrpProgluLeu.....HispheCysleuArgLeu... 93
    |||||
1772 CTTGGACTTGGTGCCACACTTAACCTGATGATGTGCTGGCGCT 1821
    |||||
94  .....leasnProAlaVally 99
1822 ACTCTTAGTCCCTGGGATATCCCGGAGACCGATGCTGCTCTAAGC 1871
    |||||
99  sserThrProasn..... 104
1872 ACTGAGACTCCGAGACTAGACTATACCGCTTATAGCAATCACCT 1921
    |||||
105  .....LeuValArgSer 108
1922 GGGAGCATAGAGGTGACGTAACCTCGGCTAGAGACTAGAAATGGCAAGC 1971
    |||||
109  TyrGlyPheTyrCysCysPro.....GlyThrGluGlyGly 121
    |||||
1972 TCTACCTTCTAGTATGTGCCCGGATGGCGGACCCCTTCAGAAAGCTAG 2021
    |||||
121  sTyrCysGlySerGlySerPheCysArgArgTrpSerCysValT 138
    |||||
2022 AAGTGCGGGGGGCTAGAAATCCCTATACTGTAAGAATGGGATTGTGAGA 2071
    |||||
138  hrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgVal 154
    |||||
2072 CCACGGGACCGGTTATTGG...CTATTAATCTCTCAAGACCTCATATA 2118
    |||||
155  LysPheSerPheValAsnSerGlyProGlyLysTyrLysMetLysLe 171
    |||||
2119 ACTGTAATGAGGACCAAAATAGCAATGAGACTCAAAATTTCAACAGTG 2168
    |||||
171  uTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLysIle 188
    |||||
2169 TCACAGACCGGCTGTGTAAACCC.....CTTAATAATAG 2203
    |||||
188  erPheThrGluLysGlyLysGlnLysnIleGlnLysTrpIleasnGly 204
    |||||
2204 ATTTCACAGACAAAGAAATTTATCC.....AAGACTGATGATACGGGA 2247
    |||||
205  MetSerTrpGlyIleValPheTyrLysTyrGlyGlyAlaGlySerTh 221
    |||||
2248 AAAACCTGGGATTAGATTCTAT...GTGTGTGACATCCAGCGCTGACA 2294
    |||||
221  rLeuThrIleArgLeuArgIleGluThrGlyThrGluProProAlaIle 238
    |||||
2295 GTTCACCACTTGCTTAATAATCACCC.....AACATGCCAGCTGTGGCAG 2338
    |||||
238  etGlyProAspLysValLeuAlaGlnGlnGlyProProAlaLeuGlu... 253
    |||||
2339 TAGGCTCGACCTGCTCTGTGTGACACAGACCTCTAGACCTCCCTC 2388
    |||||
254  .....ProProHisAsnLeuProValProGlnLeuTh 264
    |||||
2389 GCTCTCCACCTCTCTCTTCCCAAGGAGCGCCACCGCATCTCC... 2436
    |||||
264  rSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrGly 281
    |||||
2437 .....CCCGACTTAACCTCCACAGCCCTGG 2461
    |||||
281  euIleProThrAsnThrProArg.....Asn 289
    |||||
2462 CGACTAGTGCAACAACCTCCACGGTGAAGAAAACATTGTACCTTAAC 2511
    |||||
290  SerProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleG 306
    |||||
2512 ACTCG.....CTCCACACAGCGGACGAGACTTTTGTATCTTTGCA 2555
    |||||
306  nGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSer 323
    |||||
2556 GGGGGCTTCTTAACCTTAATGCTTACCAACCCAGGGGCCACTGAGCTT 2605
    |||||

323  yStrPleuCysLeuSerSerGlyProProTyrTyrGlnGlyMetAlaLys 339
    |||||
2606 GCTGGCTTGTGGCCATGAGGCCCTTATATAGCAAAATAGGCTCA 2655
    |||||
340  GluArgLysPheAsnValThrLysGlnHisArgAsnGlnCysThrTrpG 356
    |||||
2656 TCAGAGAGAGGTCCCTACTCCACCGAC...CTTGACCGGTGCGCTGGG 2702
    |||||
356  ySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCys 373
    |||||
2703 GACCAAGAGAAAGCTCACCTCTAGTAGGCTCTAGAGACAGGGTGTGCA 2752
    |||||
373  leGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrVal 389
    |||||
2753 TAGAAGAGTGGCTTACCCATCAGATCTGTGCATACAGACCTATGCC 2802
    |||||
390  TyrGlnGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArg 406
    |||||
2803 ATCAATTCCTCCGAGACATCAGTATGCTCCCTCCCAACCATAGCTG 2852
    |||||
406  pTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPhe 423
    |||||
2853 GTGGGCTTGACAGCACTGGCTCACCCCTTGCTCTCCACTTCAGTTTTR 2902
    |||||
423  snGlnSerLysAspLeuCysValMetValGlnIleValProArgValTyr 439
    |||||
2903 ATCAGACTAGAGATTCTGTATCCAGGTCCAGTGATCTCTCCACTCAT 2952
    |||||
440  TyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsn 456
    |||||
2953 TACTATCTCTGACAGACTTGTGTACAGGCTTATGACAAATTCACCCAG 3002
    |||||
456  gProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLe 473
    |||||
3003 GACTAAAGAGAGGCTGTCTACTTACCTACCTGTTTACTGGGGTGG 3052
    |||||
473  1YThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyPro 489
    |||||
3053 GAATTCACGGGGGAGATAGTACTGTTCAACTGCTTAATTAAGAGACCT 3102
    |||||
490  GlnGlnLeuGluLysGlyLeuGlyLysLeuHisAlaIleMetThrGlu 506
    |||||
3103 ATGAGCTCTCCAGCAAGCCGACAGACCTCCAGATGCCATAGATCTGA 3152
    |||||
506  pleuArgAlaLeuLysGluSerValSerAsnLeuGlnGlnSerLeuTh 523
    |||||
3153 CTTCCGGGCTCCCAAGACTCAGTCAAGCAATTAAGAGACTCACTACTT 3202
    |||||
523  erLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 539
    |||||
3203 CCTGTCCGAGGTAGTGTCCAAATAGGAGAGGCTTACTGTCTGCTT 3252
    |||||
540  LeuArgGlnGlyLysLeuCysAlaAlaLeuLysGlnGlyCysSerPhe 556
    |||||
3253 CTAAAGAAAGGTGGCTCTGTGGCGCCCTAAAGAAAGAGATGCTGTTT 3302
    |||||
556  rValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLys 573
    |||||
3303 CATAGACCACTCAGGTGACGTACGGGACTTCATGAAAAAAGCTCAAGAA 3352
    |||||
573  yLeuGlnuArgArgArgGluArgGluAlaAspGlnGlyTrpPheGlu 589
    |||||
3353 AACCTGATTAAGAGACGTAGAGCGCCACAAAAAGCAAAACGTATGAA 3402
    |||||
590  GlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeu 606
    |||||
3403 GGATGGTTCAATTAACCTCTGTGCTCACTACCTGCTATCAACATGCG 3452
    |||||
606  rGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeu 623
    |||||
3453 TGGGCCCTTAATTACTCTCTCTGTGTCTATCCCTGGGGCATGATGATCA 3502
    |||||
623  leasnArgPheValAlaPheValArgGluArgValSerAlaValGlnIle 639
    |||||

```


[illegible]

```

7929 TCTGCTCTTAGACAA 79444
seq_name: /cgn2.6/ptodata/2/1na/6A_COMB.seq:US-09-075-272-1
seq_documentation_block:
: Sequence 1, Application US/09075272
: Patent No. 6136598
GENERAL INFORMATION:
APPLICANT: MILLER, A. DUSTY
APPLICANT: WOLGAMOT, GREG
APPLICANT: BONHAM, LYNN
TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
TITLE OF INVENTION: PACKAGING CELL LINES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,272
FILING DATE: 08-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,140
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Poor, Brian W.
REGISTRATION NUMBER: 32,928
REFERENCE/DOCKET NUMBER: 14538A-003710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 467-9600
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8655 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-075-272-1

alignment_scores:
Quality: 1478.50 Length: 728
Ratio: 3.080 Gaps: 24
Percent Similarity: 65.934 Percent Identity: 44.368

alignment_block:
US-09-171-553B-10 x US-09-075-272-1 ..
Align seg 1/1 to: US-09-075-272-1 from: 1 to: 8655

6 SerTPRARGHis LeuProThrArgGlyGluProLysArgLeuArgI 22
||||| ||| ||| ||| ||| ||| :::::
5711 TCTGTGATCCAGGCTTCACACGTCAGAGAGGCCCCAGTCACATGAGTA 5760
||||| ..:::
22 LePro.....LeuSer 25
|||
5761 AACCCACGAGACACTGGGCTGTGACAGCCACTGATACCCCTTAAAG 5810
|||||
26 PAAAlaSerIleAlaTrpPheLeuThrIleThrProGlnAla 42
||||| ::||| :::::|:::
5811 TTCGGTGGTGGCAGAGAGAGCCCTCCATCACCCTGAGC..... 5849
42 IserLysAlaGluIleAspSerSerAsnProHisArgProLeuSerProT 59

```

```

5850 .....CTAGGAACACACACCTCATGCCCCAGTTCACAGT 5886
59 htrpLeuIleIleasprAspThrGlyValThrValAsnSerThrArg 75
5887 CTTGGAGATGCTTAAT...GAAAGGAGAGAGTGTATGGTACCACCT 5933
76 GlyValAlaProArgGlyThrTrpProGluLeu.....HisPheCys 90
5934 GCAGTCATACCCCTTGGACTGTGGCTGATCTCACACCTGACATCTG 5983
90 s.....LeuArgLys 93
5984 TAAATTAGCAGAGATCTCCCAATTGGACCTCCCATCATACTAGCT 6033
93 euIleAsn...ProAlaValLysSerThrProProAsnLeuVal..... 106
6034 TGAATTAACCCACCTCTGTACAAAGTGTGTCCCAACGGGTAGAGAAC 6083
107 .....ArgSerTyrG1 110
6084 ACTACTGATGTTCCGGGCGAGTCTATGAGCTAATCTTAGGGCTCACA 6133
110 yPheTyrCysCysProGlyThr.....GluLysGluLysTyrC 123
6134 GTTTATATGTTGCTCCGTGTCAGGTCAGAAAGGAAAGCTCCACAAAGAT 6183
123 ySGLyGLySerGlyLysSerPheCysArgArgTrpSerCysValThrSer 139
6184 GTAGAGGGGCATCAGACTACTTTTCCGGTAATGAGACATGTAAACACA 6233
140 AsnAspGlyAspTrpLys.....TrpProIle..... 148
6234 GGGGAAGCTTACTGGAAGCCTTCCGCTGATTGGACATCATCCGTAA 6283
149 .....SerLeuGlnAspArgValLysPheS 157
6284 ACGTGTAGTGTATGATAGCCAAACAGAGAAAGAAACCATACA 6333
157 erPheValAsnSerGly.....ProGlyLys 165
6334 AATACCTAGATCTGGGTGCTCTTAAATAATTACAGCCCCACAGACCA 6383
166 TyrLysMetLysLeuTyrLysAspLysSerCysSerProSerAspLe 182
6384 .....TGCAAGGTAAATACTGCAACCC..... 6407
182 uAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleG 199
6408 .....CTACTCATTAATAATTCAGAGAAAGGAAACAGCTCCCTG. 6449
199 IuLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrGly 215
6450 ..AGTTGGCTTAAAGGAATAGTGGGTGGCGAGTATATATTCACATA 6497
216 GlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluThrGlyTh 232
6498 AGAGACCTTGGGTTATCTTACGATTAGACTGACAGTACA.....GA 6541
232 rGluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyP 249
6542 CCTGGCAGTATCATCTAGGAGCCACAAAGTCTTACGGAACAGGCC 6591
249 rOProAlaLeuGluProProHisAsnLeuPro...ValProGluLeuThr 264
6592 CCCCACTTGCACCGGCTCCCGAGAGTCCACGCGTGCACACTCCACCA 6641
265 SerLeuArgProAspIleThrGlnProProSerAsnSerThrThrGlyLe 281
6642 ACTTCACGGGCC...TACACAGTAGAGACCCCTATTAGACAAACCTTAC 6688
281 uIleProThrAsnThrProArgAsnSerProGlyValProValLysThrG 298
5889 CTCGCCACCA.....CTCTAGATACAG 6711
298 LysGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSer 314
6712 AAACCCCTGCGCAGCTAGTTCAGGAGCCCTTTTATTAGTTAAATAG 6761
315 ThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyP 331
6762 ACTATCTTATATGACTCAATCATGTGGTATAGCTATGCTCTTAAACC 6811
331 oProTyrGlyGluGlyMetAlaLysGluArgLysPheAsnValThrLysG 348
6812 CCTTATTATGAGAGATGTGCTCAGACTACAGACTTCAATATATCTTAC 6861
348 LuHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThr 364
6862 ATCAT...TCTCAATGCTTTGGGAGAGAAAGAGAGTGTAGCTTCACA 6908
365 GluValSerGlyLysGlyThrCysIleGlyLysAlaProProSerHisG1 381
6909 GCAGTTTCAGAAATGGGCTTTGTTTAGGTGACAGTCCCGCAGATTAATG 6958
381 nHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnT 398
6959 GCACCTCTGTACACAGACTCAAAATATCCGACCTAACAAAGGTGTACGT 7008
398 yrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThr 414
7009 ATCTAGTGCCTCCCATAGACACAGTATGGGCTTCATATACAGTCTCAC 7058
415 ProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysValMe 431
7059 CCTGTATATCTATGCTGTGTTCAATAGCTCCAAAGATTTGCTATTTT 7108
431 tValGlnIleValProArgValTyrTyrHisProGlnGluValValLeu 448
7109 AGTTCACCTATTTCTAGACTCCGTGATCATGATGATAGCTCATTTCTAG 7158
448 spGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeu 464
7159 ACAATTTGAACATCGGGTC...CGCTGAAAGAGAAACCATTAATTCTT 7205
465 ThrLeuAlaValMetLeuGlyLeuGly...ThrAlaValGlyValGlyTh 480
7206 ACGTTACAGAGCTTTAGAGTTGGAGTGGCGCCAGCTGGGTAGGTAC 7255
480 rGlyThrAlaIleLeuIleThrGlyProGlnGlnLeuGlyLeuGlyLeuG 497
7256 TGGAAACCGCTGCTTAATCCAGACCCCGCATACTTTGAG..... 7295
497 LysGluLeuHisAlaIleMetThrGluAspLeuArgAlaLeuLysGluSer 513
7296 ..GAATTACGACAGATATGATAGTATGATGTCAGAGCTATAGAACA 7343
514 ValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuG1 530
7344 ATAACCAAACTAGAGAAAGATTTTAATCTCTGTCTGAGGTAGTACGCA 7393
530 nAsnArgArgGlyLeuAspLeuPheLeuArgGluGlyLeuCysA 547
7394 AAATAGGAGAGATTTAGATTTGTTATTCTTAAAGAAAGAGACTTTGTG 7443
547 IAlaLeuLysGluGluCysCysPheTyrValAsnHisSerGlyAlaIle 563
7444 CTGCTCCAAAGAGAGATTTGTTTATGTCGACCACTCGGGAGTCATC 7493
564 ArgAspSerLeuAsnLysLeuArgLysLysLeuGluArgArgArgG1 580
7494 AAACATTCTATGGCCAAACTAGAGAACGCCCTGATTTAGCTCAAAAGA 7543
580 uArgGluAlaAspGlnGlyTrpPheGluGluGlyTrpPheAsnArgSerPro 597
7544 GAGAGAGAGCAACAGAGATGTTTGAAGCTGTTTAATATAGTCCCTT 7593

```

[illegible]

```

6636 CCCCTACCC.....AAACCTGCCAAGTCTCCCGCCGCTC 6670
275 rAnSerThrThrGlyLeu...ProThrAsnThrProArgAsnSer 291
||:||||| ||||| ||||| ||||| ||||| |||||
6671 TAGTTCGACTCCACATTGATTCCCGTCCCGCTCCACCTCCACTAGCCCC 6720
291 rGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGly 307
||:||||| ||||| ||||| ||||| ||||| |||||
6721 CG.....CCAGCAGGAACGGGAGACAGATTACTAATCTAGTACAGGGA 6764
308 AlaHeglnAlaIleAsnSerThrAspProAspAlaThrSerSerGly 324
||:||||| ||||| ||||| ||||| ||||| |||||
6765 GCTTACCGAGGACCTTACCTTACCACTGATATAAATCAAGAGTCTG 6814
324 PLeuCysLeuSerSerGlyProProTyrTyrGlyGlnGlyMetAlaLysGln 341
||:||||| ||||| ||||| ||||| ||||| |||||
6815 GTTATGCTTACTGTCTGACCCCTTATAGAGGGGCTTGCCTCTAG 6864
341 rGlySpheAsnValThrLysGlnHisArgAsnGlnCysThrTrpGlySer 357
||:||||| ||||| ||||| ||||| ||||| |||||
6865 GTACTTATTCACACCATACCTCTCCCGAGCTTACTGCTCCGTGCTCC 6914
358 ArgAsnLysLeuThrLeuThrGlnValSerGlyLysGlyThrGlyIleG 374
||:||||| ||||| ||||| ||||| ||||| |||||
6915 CAACACACAGCTGACCTGCTCCGAGAGTACTGACGGGACTCTGCATAGG 6964
374 YLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrG 391
||:||||| ||||| ||||| ||||| ||||| |||||
6965 AACAGTCCCAAAACTCACCAGGCGCTGTGCACACATACCTT.....A 7008
391 LglnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTyr 407
||:||||| ||||| ||||| ||||| ||||| |||||
7009 AGGAGAGGAAAGGCTTACTATCTACTATGCCCCCAGGAAACATATGTG 7058
408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnG 424
||:||||| ||||| ||||| ||||| ||||| |||||
7059 GCATGTATACACTGACCTACTCTCCATGCTTGCACCGTCTTAATCG 7108
424 nSerLysAspLeuCysValMetValGlnIleValProArgValTyrTyr 441
||:||||| ||||| ||||| ||||| ||||| |||||
7109 CACACGTACTATTTGCGTTTCGGAATATATGCCCCAGGCTACCTAC 7158
441 lSPROGluGlnValValLeuAspLysTyrArgTyrAsnArgPro 457
||:||||| ||||| ||||| ||||| ||||| |||||
7159 ATCTCCCACTTACGCTATATGCCAGTTTGAA...AAATCCCATAGACAT 7205
458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGly..... 471
||:||||| ||||| ||||| ||||| ||||| |||||
7206 AAAAGAGAACAGGTCTCTTACCTTGCCCTTATTTAGGTGGCTAAC 7255
472 ...LeuGlyThrAlaValGlyValGlyThrAlaAlaLeuIleT 487
||:||||| ||||| ||||| ||||| ||||| |||||
7256 TATGGGTGGCATGCGCGGGAGTAGGAGACGAGAACCTACCGCCCTG 7304
487 hrGlyProGlnGlnLeuGlnLysGlyLeuGlnLysAlaAlaMet 503
||:||||| ||||| ||||| ||||| ||||| |||||
7305 ..GCCACCCACAGTTTACG.....CAGCTCCATGCTGCCGCTA 7340
504 ThrGluAspLeuArgAlaLeuLysGlnSerValSerAsnLeuGlnLys 520
||:||||| ||||| ||||| ||||| ||||| |||||
7341 CAAGATGATCTCAAAAGAGTCGAAAGTCATTAACCTAGAGAAAGTC 7390
520 rLeuThrSerLeuSerGlnValValLeuGlnAsnArgArgGlyLeuAsp 537
||:||||| ||||| ||||| ||||| ||||| |||||
7391 TCTTACTCTGTTGCTGAGGTGTAGTCGCAAAATCGACAGAGCCTAGACC 7440
537 euleuPheLeuArgGlnGlyLysCysAlaAlaLeuLysGlnGlyCys 553
||:||||| ||||| ||||| ||||| ||||| |||||
7441 TGTGTCTCTTAAAGAGGAGGACTGTCTGCTGCCCTTAAAGAAAGATGT 7490
554 CysPheTyrValAspHisSerGlyAlaAlaLeuArgAspSerMetAsnLys 570
||:||||| ||||| ||||| ||||| ||||| |||||
7491 TGTTCATGCTGACCATACAGGCGCTAGTAGATAGTATGCGCAATAT 7540

```

```

570 uArgLysLysLeuGluArgArgArgGlnuArgGlnAlaAspGlnGlyT 587
||||| ||||| ||||| ||||| ||||| |||||
7541 AAGAGAGACACTCTCTCAGACAGCAAAACTATTGTAGTGCAGCAAGAT 7590
587 rPheGlnGlyTyrPheAsnArgSerProTrpMetThrThrLeuLeuSer 603
||||| ||||| ||||| ||||| ||||| |||||
7591 GGTTCGAGAGAGGTGTTAACAGATCCCTCGTTTACCACTGATATCC 7640
604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGlyPyr 620
||||| ||||| ||||| ||||| ||||| |||||
7641 ACCATATAGGGGCTCTCATTAATCTCTACTTAATTTCTTTTGGACC 7690
620 oCysLeuIleAsnArgPheValAlaPheValArgGlnuArgValSerAla 637
||||| ||||| ||||| ||||| ||||| |||||
7691 CTCATTCCTTAATCGATTAGTTCAATTTGTAAAGACAGATCTCAGTAG 7740
637 aLglnIleMetValLeuArgGlnGlnGlnTyrGlnGlyLeu 649
||||| ||||| ||||| ||||| ||||| |||||
7741 TCCAGGCTTATGCTCTGACTCAACATATACCACTGACTA 7778

```

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8
seq_documentation_block:
; Sequence 8, Application US/08886642
; Patent No. 5952474
; GENERAL INFORMATION:
; APPLICANT: Pfizer, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 45 Rockefeller Plaza, Suite 2800
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/886,642
; FILING DATE: 01-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/110,300
; FILING DATE: 20-AUG-1993
; APPLICATION NUMBER: 07/938,100
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hone, William J.
; REGISTRATION NUMBER: 26,739
; REFERENCE/DOCKET NUMBER: 07763/010002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212/765-5070
; TELEFAX: 212/258-2291
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-886-642-8

```

```

alignment_scores:
Quality: 1329.00      Length: 696
Ratio: 2.973          Gaps: 23
Percent Similarity: 64.224      Percent Identity: 42.816

```

```
alignment_block:
    rc 00 171 f33
```

US-09-171-553B-10 x US-08-886-642-8

Align seg 1/1 to: US-08-886-642-8 from: 1 to: 8323

[illegible]

```

|||||
7441 TGTGTTCTTAAAGAGGAGGAGCTGTGTCTGCTTAAAGAGAGATGT
554 CysPheTyrValAspHisSerGlyAlaIleArgAspSerMetLysLysLe 570
|||||
7491 TGTTCATCTGCTGACCATACAGGCGCTAGTAAGATAGTATGGCCAAAT 7540
570 uArgLysLysLeuGluArgArgArgGluArgGluAlaAspGlnGlyT 587
|||||
7541 AAGAGAGAGACTCTCTCAGAGCAAAACTATTGAGTCAAGCCAAAGAT 7590
587 rPheGlnGlyTyrPheAsnArgSerProTrpMetThrThrLeuSer 603
|||||
7591 GGTTCGAGGAGATGTTTAACAGATCCCGCTGTACACAGTGTATATCC 7640
604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGlyPr 620
|||||
7641 ACCATCATGGGCGCTCTCATATACCTCTACTAATCTCGCTTTTGAGCC 7690
620 OCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerLay 637
|||||
7691 CTGCATCTTATATGATAGTCAATTTGTTAAAGACAGATCTCAGTAG 7740
637 aGlnIleMetValLeuArgGlnGlnGlnGlnGlyLeu 649
|||||
7741 TCCAGGCTTATGCTCTGACTCAACAATACCCAGCTA 7778

```

seq_name: /cgn2_6/prodata/2/ina/PC/US93_COMB.seq: PCT-US93-08041-8

```

seq documentation block:
; Sequence 8, Application PC/TUS9308041
; GENERAL INFORMATION:
; APPLICANT: The Public Health Research Institute of the City of
; APPLICANT: New York, Inc.
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull and Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08041
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,100
; FILING DATE: 28-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 20-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H.
; REGISTRATION NUMBER: 32,140
; REFERENCE/DOCKET NUMBER: 11698A50
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

```

```

; ANTI-SENSE: NO
PCT-US93-08041-8
alignment_scores:
  Quality: 1329.00      Length: 696
  Ratio: 2.973          Gaps: 23
  Percent Similarity: 64.224  Percent Identity: 42.816
alignment_block:
US-09-171-553B-10 x PCT-US93-08041-8
Align seg 1/1 to: PCT-US93-08041-8 from: 1 to: 8323
16 GluProLysArgLeuArgGlyLeuProLeuSerPheIleAlaThrPhe 32
5817 GACCCGCGGAGCCTCTTAATCCCTTAATTCCTC.....TT 5851
32 eLeuThrLeuThrIleThrProGlnAlaSerLysArgLeuIleAsp 49
5852 CCTGTCTCTC.....AAGGGCCAGATCCGAGCACC 5886
49 eSerAsnProHisArgProLeuSerProThrTrpLeuIleAspPro 65
5887 GCTCCAGCCCTCACCAGTCTACACATTAATCTGGAGAGTCAATGG 5936
66 AspThrGlyValThrValAsnSerThrArgGlyValAlaProArgGly 82
5937 GATCGG...GAGACAGTATGGCAATATCAGGCAACCACCTCTGTGAC 5983
82 rTrpTrpProGluLeu.....CysLeuArg 87
5984 TTGTGTGCCAGTCTCTACCCCAATTTGTGTATGTAGCTTCAGTGGGC 6033
88 .....HisPhe.....CysLeuArg 89
6034 CGCCCACTGGGGCTAGAGTATCAGCCCCCTATCTCGCCCGGGG 6083
90 .....CysLeuArg 93
6084 CCCCCTTTGTTGTCAGAGGAGCAGCGGAAACGTTGACGCTGTGCCAGAA 6133
93 uIleAsnProAlaValLysSerThrPro.....A 104
6134 CTGCAAGAGCCCTTGACCTCCCTCAACCCCTGCTGCACACTGCTGGA 6183
104 snLeuValArg.....SerTyrGlyPheTyr 112
6184 ACAACTTAAGCTGACACCAAGTAATCAATAATCAAGTGAGGATTTAT 6233
113 CysCysProGlyThrGluLys.....GluLysTyrCysGlyLys 126
6234 GTCTGCCCGGGTACATCGCCCCCGGAGAACCAAGTCTGTGGGGTCC 6283
126 rGlyGluSerPheCysArgArgTrpSerCysValThrSerAsnAspGly 143
6284 AGACTCTTCTACGTGCTCTTGGGGCTGCGAGCAACCGTAGAGTAT 6333
143 sPrrLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
6334 ACTGGAG...CCCTCTCTTCTTGGGACATCAACA.....GTA 6371
160 AsnSerGlyProGlyLysTyrLysMetLysLeuTyrLysAsp...Lys 175
6372 GACACACATCTCACCTTAACGAGGCTGTACAGTATGCAAAACATATA 6421
175 sSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGlu 192
6422 GTGGTGCATGCC.....TTGGCTATCCGGTTTAAACAACG 6456
192 ysgLysLysGlnGluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGly 208
6457 CCGGGAACAG.....GTCACCTCATGACACTGACACTATATGGGGT 6500

```

209 IleValPheTyrTyrGlyGlyAlaGlySerThrLeuThrIleAr 225
 225 gIeuArGlyIleGluThrGlyThrGluProProValAlaMetGlyProAspL 242
 6551 ACTCACTATTATCAA...AACTAGAGACCTGGATCCCAATAGAGACCAAAAC 6597
 242 ySValLeuAlaGluGlyProProAlaLeuGluProProHisAsnLeu 258
 6598 CCGTCCTGGAGACCAA...CTTTCCGTTCCCGCTACTTAT 6635
 259 ProValProGluLeuThrSerLeuArgProAspIleThrGluProProse 275
 6636 CCGCTACCC...AAACCTGCCAAGTCTCCCGCGCTC 6670
 275 rAsnSerThrThrGlyLeuLeu...ProHisThrProArgAsnSerP 291
 6671 TAGTTGAGCTCCCACTGATATTTCCCGCCCACTCCCACTCCAGCCGCC 6720
 291 roGlyValProValLysThrGlyGluArgLeuPheSerLeuIleGly 307
 6721 CG...CCAGAGGAGAGGAGAGAGATTAATCTACTACAGGGA 6764
 308 AlaPheGluAlaIleAsnSerThrAspProAspAlaThrSerSerCysTr 324
 6765 GCTTACCAGGAGCTCAACCTGACCAACCTGATAAACCTCAAGAGTGTG 6814
 324 pIeuCysLeuSerSerGlyProProGlyTyrGluGlyMetAlaLysGluA 341
 6815 GTTATGCTCTAGTGTCTGGACCCCTATATAGAGGGGTTGCGCTAG 6864
 341 rGlySPheAsnValThrLysGluHisArgAsnGlnCysThrTrpLysSer 357
 6865 GATCTTATTCACACCATACCTGTGCCCGCACTACTGCTCCGCGCTCC 6914
 358 ArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleG 374
 6915 CAACACAAAGCTGACCTGTCCGAAGTGAAGTGAAGGAGCTGTGATAG 6964
 374 yLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrG 391
 6965 AACAGTCCCAAAACTCACAGGCTGTGCAACACTACCTT...A 7008
 391 IuGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrp 407
 7009 AGGCAAGGCAAGGCTCTACTATCTAGTTGCCCCACAGGAGCTATGTGG 7058
 408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnG 424
 7059 GCATGTACACTGAGACTCATCCTACCTATGTCACCGTGGCTTAATCG 7108
 424 nSerLysAspLeuCysValMetValGlnIleValProArgValTyrTrH 441
 7109 CACCACTGACTATGCTGTCTGTGGATTAAGGCCAGGGCTCACCTTCC 7158
 441 iSProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPro 457
 7159 ATCTCCCAAGTATAGCTATAGCCAGTTTGA...AAATCCCATAGACAT 7205
 458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGly... 471
 7206 AAAAGAGAACAGTGTCTTAACCTTGGCTTATATAGTGGGCTAAC 7255
 472 ...LeuGlyThrAlaValAlaGlyValGlyThrGlyThrAlaLeuIleT 487
 7256 TATGGGTGGCATCGCGGAGAGTGGAGGAGAACTACCGCCCTGGGC. 7304
 487 ThrGlyProGlnGlnLeuGluLysGlyLeuGlyLysLeuHisAlaLam 503
 7305 ...GCCACCCAGAGTTTCAG...CAGCTCCATGCTGCCGTA 7340

504 ThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluLys 520
 7341 CAAGATGATCTCAAGAAAGTCGAAAAAGTCAATTACTTACCAAGAAAGTCC 7390
 520 rLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspL 537
 7391 TCTTACTTCTGTTCTGAGGTGTACTGTCAAGAAATCGACGAGGCTTAGACC 7440
 537 eulePheLeuArgGluGlyLeuCysAlaAlaLeuLysGluLys 553
 7441 TGTTCCTTCTTAAAGAGGAGGAGCTGTGTGCTGCCCTTAAAGAAAGTGT 7490
 554 CysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLe 570
 7491 TGTTCCTTCTTAAAGAGGAGGAGCTGTGTGCTGCCCTTAAAGAAAGTGT 7540
 570 uArgLysLysLeuGluValArgArgArgGluArgGluAlaAspGlnGlyT 587
 7541 AAGAGAGAGCTCTCTCAGAGACAAAACTATTGAGTCGAGCCAGGAT 7590
 587 rPheGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSer 603
 7591 GGTTCGAAGGATGTGTTTACAGATCCCTGTGCTTACAGCTGATATTC 7640
 604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGlyPr 620
 7641 ACCATCATGAGGCGCTTCATATATACCTCTACTTATTTGCGCTTTTGAGCC 7690
 620 cCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlav 637
 7691 CTGCATTCCTTATGATGATTAGTTCAATTGTTAAAGACAGAGTCTCAGTAG 7740
 637 alGlnIleMetValLeuArgGlnGlnGlnGlnGlyLeu 649
 7741 TCCAGGCTTATGCTCTGACTACACATACACAGCTTA 7778

seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:us-09-011-745-6

seq_documentation block:

? Sequence 6: Application US/09011745
 ? Patent No. 6165715
 ? GENERAL INFORMATION:
 ? APPLICANT: Collins, Mary KL
 ? APPLICANT: Weiss, Robin A
 ? APPLICANT: Takeuchi, Yasuhiro
 ? APPLICANT: Cosset, Francois-Loic
 ? TITLE OF INVENTION: Expression systems
 ? FILE REFERENCE: 09/011,745
 ? CURRENT APPLICATION NUMBER: US/09/011,745
 ? EARLIER FILING DATE: 1998-06-22
 ? EARLIER APPLICATION NUMBER: PCT/GB96/02061
 ? EARLIER FILING DATE: 1996-08-23
 ? EARLIER APPLICATION NUMBER: GB9517263.1
 ? NUMBER OF SEQ ID NOS: 29
 ? SOFTWARE: PatentIn Ver. 2.0
 ? SEQ ID NO 6
 ? LENGTH: 6061
 ? TYPE: DNA
 ? ORGANISM: Artificial Sequence
 ? FEATURE:
 ? OTHER INFORMATION: Description of Artificial Sequence: Portion of
 ? FEATURE:
 ? NAME/KEY: misc_feature
 ? LOCATION: (3807)
 ? OTHER INFORMATION: n is any nucleotide
 ? FEATURE:
 ? NAME/KEY: misc_feature
 ? LOCATION: (3808)
 ? OTHER INFORMATION: n is any nucleotide
 ? FEATURE:
 ? NAME/KEY: misc_feature


```

1 LOCATION: (3809)
2 OTHER INFORMATION: n is any nucleotide
3
4 FEATURE:
5 NAME/KEY: misc_feature
6 LOCATION: (3810)
7 OTHER INFORMATION: n is any nucleotide
8
9 FEATURE:
10 NAME/KEY: misc_feature
11 LOCATION: (3995)
12 OTHER INFORMATION: n is any nucleotide
13
14 FEATURE:
15 NAME/KEY: misc_feature
16 LOCATION: (3996)
17 OTHER INFORMATION: n is any nucleotide
18
19 FEATURE:
20 NAME/KEY: misc_feature
21 LOCATION: (3997)
22 OTHER INFORMATION: n is any nucleotide
23
24 FEATURE:
25 NAME/KEY: misc_feature
26 LOCATION: (3998)
27 OTHER INFORMATION: n is any nucleotide
28
29 OS-09-011-745-6

```

```
alignment_scores:
  Quality: 1284.00      Length: 734
  Ratio: 2.828          Gaps: 26
Percent Similarity: 61.853  Percent Identity: 41.008
```

```
alignment_block:
  00 131 f33d 10 00 00 011 7ae 6
```

Align seg 1/1 to: US-09-011-745-6 from: 1 to: 6061

```

2 HispHrThLeuSerTyrPArgHisLeuProHrArgGlyGlyGlyGlyGlyGlyGly 18
||| ::::::::::: ||| ::|
1186 CATTGAGCGTTGGATACACAGCCG...CCAGTGAAGGCTCCGACGCC 1352
18 SarGLeuArgIleProLeuSerPheAlaSerIleAlaIlePheLeuHrL 35
: ::||| ::|
1333 G.....GGGTGGACATCCCT 1349
35 eutHrIleThrProGlnAsr...SerLySArGLeuIleAspSerSer 50
||| ::::::::::: ||| ::|
1350 AGACTGACATGGCGGGTTCAACGCTCTCAAAACCCCTTAAATAAAGATT 1399
51 AsnProHisAspProLeuSerPro..... 58
||||| ::| ||| |||
1400 AACCCGCGAGGCCCTTAATCCCTTAATTTCTTGATGCTCAGAGGGGT 1449
59 .....Thr 60
1450 CAGTACTGCTGGCCCGGCTCCAGTCACGTCATCAATCAATATATATCACT 1499
||| |||
60 rPLeuIleIleAspProAspThrGlyValThrValAsnSerThrArgGly 76
|| ::| ::| ||| ||| ::| |||
1500 GGGAGGTACAACATGAGATCG...GAGAGGATGTGGCAACTTGTGCG 1546
77 ValAlaProArgGlyThrTTPTrProGluLeu..... 87
||| ||| ||| ||| ||| |||
1547 AACCAACCTCTGTGGACGCTGTGGGCTTACCTACCCAGATTATGTAT 1596
88 .....HisPheCysLeuArgLeuIleAsnPro 97
1597 GTTAGCCCAACATGACCATCTTATTGGGGCTAGAAATATCAATCCCTT 1646
97 laValLysSerThrProPro..... 103
||| ::| |||
1647 TT.....TCTTCTCCCGCGGGGCCCTGTGTGCTCAGGGGGCAGCAGC 1690
103 ..... 103

```

1691 CAGAGCTGTTCCAGAGACTGCGGCAAGAACCTTTAAACCTCCCTCACCCCTCG 1744

104AsnLeuValArg..... 107

1741 GTGCACAACACTGCTCGGACACAGACTAACCTAGACCAGACAATCATATAAT 1790

108 ..SerThrGlyPheThrCysCysProGlyThrGluIys.....Glu 120

1791 CAATGAGGAGATTATATATTGTCGCCCGGGCCGACCGGCCCGGAGAAATCC 1840

121 LysTyrCysGlyLeuIleSerGlyIleuSerPheCysArgArgThrPheSerGlyVal 137

1841 AAGCTCATGTGGGGGTCCACAGACCTCTTCACGTGCTGCTATTGGGGCTGTGA 1890

137 LTrpSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgVal 154

1891 GACAACCGGTGAGAGCTTACTGGAGA...CCCTCCATCATCAGGAT.... 1933

154 allsPheSerPheValAsnSerGlyProGlyLysTyrLysMetMetLys 170

1934TTTCATCAACAGTAACAACAATCTACCTGTGACAGAGCTGTCCAG 1978

171 LeuTyrLysAsp...LysSerCysSerProSerAspLeuAspTyrLeuLys 186

1979 GTATGCAAGATAATATAGTGGTGCAACCC.....TTAGT 2013

186 sIleSerPheThrGlyLysGlyLysGlnGluAsnIleGlnLysTrpIleA 203

2014 TATTGCGTTTACAGACGCCGGGAGACGG.....GTACTCTCTGGACCA 2057

203 snuGlyMetSerTrpGlyLysLeuAlaPheTyrLysTyrGlyGlyValAlaGly 219

2058 CAGAGCAATTACTGCGGGCTTACGTTGTATGCTCCGGCAAGATCCAGGG 2107

220 SerThrLeuThrIleArgLeuArgIleGluThrGlyTrpGluProProVal 236

2108 CTTACATTGTGGATTCGCTACTGATATACCA...AATCATAGACCOCGGGT 2154

236 IAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeuG 253

2155 CCAATATGGGCGCAACCCGCTTGTGGACAGCCAAAGCCA.....CTCT 2198

253 LuProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAsp 269

2199 CCAAGGCCAAA.....CCTGTTAAGTCG.....CCTCA 2227

270 IleThrGlnProProSerAsnSerThrThrGlyLeuIleProThrAsnTh 286

2228 GTCCACCAACACCCACGAGTGGG.....ACTCTCTCTCCCTACCCAACT 2271

286 rProThrGlnSerProGlyValProValLysThrGlyGlnArgLeuPheS 303

2272 TCCA.....CCGGCGGGAAAGGAAATAGCTGCTGTA 2303

303 eLeuIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspProAspAla 319

2304 ACTTAGTAGAGGAGGAGGCTACCAAGCCCTCAACCTCAGCAGTCTGACAAA 2353

320 ThrSerGlySerTrpLeuCysLeuSerSerGlyProProTyrTyrGluGln 336

2354 ACCCAAGAGTGTGTTGTGTAGTAGAGGGGAGCCCTCTACTAGGAAG 2403

336 yMeAlaLysGluArgLysPheAsnValThrLysGlnHisArgGanglinc 353

2404 GGTGGCCGTCCTGGATACCTACACCAACCATCTCTGCTCCAGCAACT 2453

353 yStrThrGlySerArgAsnLysLysLeuThrLeuThrGluValSerGlyLys 369

2454 GTCGCGGAGGCTCCCAACAAGATTGACCTGTCCGAAGTAGCAGCGACAG 2503

370 GluThrCysIleGlyLysAlaProProSerHisGlnHisLeuCysTyrSe 386

2504 GGACTGTGCATAGGAGAGTGTCCCAAAACATCAAGGCGCTATGTGAATAC 2553

```
386 rThrValValTyrGlnAlaSerGluAsnGlnTyrLeuValProGlyT 403
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2554 CACCCAGACAGACAGTCCAGGCTCC.....TATTATCTAGTGGCCCTA 2597
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
403 yfAsnArgTrpTPrAlaCysAsnThrGlyLeuThrProCysValSerThr 419
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2598 CAGGTACCATGTGGCTGTGTAGTACCGGGCTTACTCCATGCATCTCCACC 2647
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
420 SerValPheAsnGlnSerLysAspLeuCysValMetValGlnIleValPr 436
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2648 ACCATACAGAACCTTACCACATGATATGTGTCTTGTGAACTCTGGCC 2697
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
436 oArValAlaTyrTrpHisProGlnIleValValLeuAspGlnTyrAspTyrA 453
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2698 AAGAGTCACCTATTCATCTCCCGCCTATGTTACGGCTGTTTGAG...A 2744
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
453 rGTrfAsnArgProLysArgGluProValSerLeuThrLeuAlaValMet 469
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2745 GATCCACCGCACACAAAGAGACCGGTGTGCTTACCCCTGGCCCTATTA 2794
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
470 LeuGly.....LeuGlyThrAlaValGlyValGlyThrGlyTh 482
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2795 TTGGGTGGACTAACCATGGGGGAATGCGCGTGAATAGGAACAGGGGAC 2844
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
482 rAlaAlaLeuIleThrGlyProGlnIleuGlnLysGlyLeuGlyGluL 499
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2845 TACTGCTCTAATG...GCCACTCAGCAATTCGAG.....CAGC 2879
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
499 eunHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSer 515
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2880 TCCAGAGCCGCACTACAGATGATCTCAGGAGAGTGAATAATCATCTCT 2929
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
516 AsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnAr 532
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2930 AACCTAAGAAAGCTCTCTCACTCCCTGCTGAACTTGTCTACAGAAATCG 2979
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
532 gArGlyLeuAspLeuPheLeuArgGlnGlyLeuCysAlaAlaL 549
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2980 AAGGGGCTTAGACTTGTATTCTTAAAGAAAGAGGGCTGTGTGCTGCTC 3029
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
549 euLysGlnGluCysCysPheTyrValAspHisSerGlyAlaIleArgAsp 565
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3030 TAAAGAAAGAAATGCTTCTCTATGGGAGCCACACAGACTAGTAGAGAC 3079
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
566 SerMetAsnLysLeuArgLysLysLeuGluArgArgArgGluArgG1 582
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3080 ACATGGGCCAAATGAGAGAGAGCTTAATCAGAGACAAACTGTGTGA 3129
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
582 uAlaAspGlnGlyTrpPheGlnGlyTyrPheAsnArgSerProTrpMetT 599
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3130 GTCACTCAAGAGATGTTTGTAGGACTGTTTAAACAGATCCCTTGTGTTTA 3179
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
599 hrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeu 615
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3180 CCACCTGTATATCTACCATATGAGAGCCCTCATTTGTACTCTATATGAT 3229
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
616 LeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgG1 632
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3230 TTGCTCTTGGACCTGCATCTTATATGATAGTCAATTGTTAAGA 3279
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
632 uArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyL 649
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3280 CAGGATCTCAGTATCCAGGCTTGTAGTCTGACTCAACATATACACAGAC 3329
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
649 eu 649
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3330 TA 3331
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 03:01:22 ; Search time 212.72 Seconds
(without alignments)
21.294 Million cell updates/sec

Title: US-09-171-553B-13

Perfect score: 20

Sequence: 1 gatgctctcctgcctt 20

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0.

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_MN:*

2: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*

3: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*

5: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	20	100.0	7333	US-08-766-528-2
2	18.4	92.0	3482	US-09-111-085-3
3	18.4	92.0	8060	US-08-766-528-1
4	18.4	92.0	8132	US-08-766-528-3
5	16.8	84.0	72928	US-09-009-913-1
6	16	80.0	23	US-08-766-528-14
7	16	80.0	23	US-08-766-528-15
8	16	80.0	2033	PCT-US93-12464-1
9	16	80.0	2400	US-08-800-2918-7
10	15.4	77.0	4060	US-08-308-949A-1
11	15.4	77.0	31571	US-08-323-443B-1
12	15.4	77.0	53526	US-08-658-136-2
13	15.4	77.0	53577	US-08-658-136-1
14	15.2	76.0	3103	US-08-826-246-3
15	15.2	76.0	3103	US-08-944-495-3
16	15.2	76.0	3103	US-09-126-640-2
17	15.2	76.0	3103	US-08-925-588-3
18	15.2	76.0	3111	US-09-487-444-3
19	15.2	76.0	8655	US-09-075-272-1
20	15.2	76.0	15101	US-08-799-464A-14
21	15.2	76.0	15101	PCT-US95-09927-14
22	15.2	76.0	15108	US-08-157-005-1
23	15.2	76.0	15108	US-08-747-863-1
24	15	75.0	68	PCT-US95-17026-8
25	15	75.0	70	PCT-US95-17026-9
26	15	75.0	3876	PCT-US95-17026-1
27	14.8	74.0	489	US-08-081-072-9

C 28	14.8	74.0	489	1	US-08-449-093A-9	Sequence 9, Appl
C 29	14.8	74.0	569	2	US-08-483-695-44	Sequence 44, Appl
C 30	14.8	74.0	569	2	US-07-965-285-44	Sequence 44, Appl
C 31	14.8	74.0	569	2	US-08-487-231-44	Sequence 44, Appl
C 32	14.8	74.0	569	4	US-09-201-912-44	Sequence 44, Appl
C 33	14.8	74.0	7886	2	US-08-751-189-2	Sequence 2, Appl
C 34	14.8	74.0	7886	4	US-09-060-836-2	Sequence 2, Appl
C 35	14.8	74.0	7886	4	US-09-184-445-2	Sequence 2, Appl
C 36	14.4	72.0	1539	2	US-08-828-596-1	Sequence 1, Appl
C 37	14.4	72.0	1539	4	US-09-360-197-5	Sequence 5, Appl
C 38	14.4	72.0	2567	3	US-08-993-260-4	Sequence 4, Appl
C 39	14.2	71.0	1160	2	US-08-955-138-5	Sequence 5, Appl
C 40	14.2	71.0	1470	6	US-08-955-138-5	Sequence 5, Appl
C 41	14.2	71.0	3083	1	US-08-346-849-1	Sequence 1, Appl
C 42	14.2	71.0	3083	2	US-08-293-284A-1	Sequence 1, Appl
C 43	14.2	71.0	4020	4	US-09-050-159-130	Sequence 130, App
C 44	14.2	71.0	4024	4	US-09-162-484-18	Sequence 18, Appl
C 45	14.2	71.0	4928	1	US-08-399-561-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF INVENTIONS: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-2
Query Match 100.0%; Score 20; DB 4; Length 7333;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 gatgctctcctgcctt 20
|||||

Db 4442 GATGCTCTCTGCTTGG 4461

RESULT 2

US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111.085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 3; Length 3482;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgctctctgaccttg 20
||| ||||| ||||| |||||

Db 528 gatagctctctgaccttg 547

RESULT 3

US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8060 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 4; Length 8060;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgctctctgaccttg 20
||| ||||| ||||| |||||

Db 7763 GATAGCTCTCTGCTTGG 7782

RESULT 4

US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-3

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 4; Length 8132;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgctctctgaccttg 20
||| ||||| ||||| |||||

Db 5237 GATAGCTCTCTGCTTGG 5256

RESULT 5

US-09-009-913-1/C
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:

APPLICANT: Axy's Pharmaceuticals, Inc.
TITLE OF INVENTION: Asthma Related Genes
NUMBER OF SEQUENCES: 339
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTED for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 72928 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match 84.0%; Score 16.8; DB 3; Length 72928;
Best Local Similarity 90.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gatgctctcctgccccttg 20
||||||| | |||||
Db 46404 GATGCTCTCCATGCCCTTTG 46385

RESULT 6
US-08-766-528-14/c
Sequence 14, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-14

Query Match 80.0%; Score 16; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gctctcctgccccttg 20
||||||| | |||||
Db 23 GCTCTCTGCCCCCTTG 8

RESULT 7
US-08-766-528-15
Sequence 15, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-15

Query Match 80.0%; Score 16; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 gctctcctgccttg 20
|||||
Db 1 GCTCTCTGCGCTTTG 16

RESULT 8
PCT-US93-12464-1/c
Sequence 1, Application PC/TUS9312464
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California, et al.
TITLE OF INVENTION: METHOD FOR INHIBITION OF CELL
ADHESION TO RECEPTORS CONTAINING SELECTINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
CITY: LOS ANGELES
STATE: CALIFORNIA
COUNTRY: US
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/12464
FILING DATE: 21-DEC-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HOWEELS, STACY L.
REGISTRATION NUMBER: 34,842
REFERENCE/DOCKET NUMBER: PD2929
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: ALPHA-L-FUCOSIDASE
FEATURE:
NAME/KEY: CDS
LOCATION: 19..1401
PCT-US93-12464-1

Query Match 80.0%; Score 16; DB 5; Length 2035;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 gctctcctgccttg 20
|||||
Db 1851 GCTCTCTGCGCTTTG 1836

RESULT 9
US-08-800-291B-7
Sequence 7, Application US/08800291B
Patent No. 6153740
GENERAL INFORMATION:
APPLICANT: J.D. Young & C.E. Cass
TITLE OF INVENTION: CDNA ENCODING NUCLEOSIDE TRANSPORTER
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,291B
FILING DATE: 13-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/499,314
FILING DATE: 7-JULY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hallie, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07254/044M01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2400 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2401
US-08-800-291B-7

Query Match 80.0%; Score 16; DB 3; Length 2400;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 gctctcctgccttg 20
|||||
Db 2256 GCTCTCTGCGCTTTG 2271

RESULT 10
US-08-308-949A-1
Sequence 1, Application US/08308949A
Patent No. 5580703
GENERAL INFORMATION:
APPLICANT: Kolin, Robert M.
APPLICANT: Berns, Kenneth I.
APPLICANT: Linden, Ralph M.
TITLE OF INVENTION: Human Adeno-Associated Virus Integration
SITE OF INVENTION: Site DNA and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,949A
FILING DATE: September 20, 1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,127
FILING DATE: September 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC92-10F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-9540
TELEFAX: 617-861-6240
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-308-949A-1

Query Match 77.0%; Score 15.4; DB 1; Length 4060;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 tggctctctgcccctt 19
|||||
Db 2970 tggctctctgcccctt 2986

RESULT 11
US-08-323-443B-1/c
Sequence 1, Application US/08323443B
Patent No. 5654170
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W.
APPLICANT: LANDES, GREGORY M.
APPLICANT: BURR, TIMOTHY C.
APPLICANT: CONNORS, TIMOTHY D.
APPLICANT: DACKOWSKI, WILLIAM R.
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,443B
FILING DATE: 12-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, S. Peter
REGISTRATION NUMBER: 25,351
REFERENCE/DOCKET NUMBER: 0372/0A462
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 31571 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: PKD1 GENOMIC
US-08-323-443B-1

Query Match 77.0%; Score 15.4; DB 1; Length 31571;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gatgctctctgcccct 17
|||||
Db 13938 GCTGGCTCTCTGCCCCCT 13922

RESULT 12
US-08-658-136-2/c
Sequence 2, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W.
APPLICANT: LANDES, GREGORY M.
APPLICANT: BURR, TIMOTHY C.
APPLICANT: CONNORS, TIMOTHY D.
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-5415
TELEFAX: 508-872-8400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 53526 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match 77.0%; Score 15.4; DB 3; Length 53526;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gatgctctctgcccct 17
|||||
Db 14923 GCTGGCTCTCTGCCCCCT 14907

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826, 246
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
Prior APPLICATION DATA:
APPLICATION NUMBER: 08/799, 910
FILING DATE: 13-FEB-1997
Prior APPLICATION DATA:
APPLICATION NUMBER: 60/011, 787
FILING DATE: 16-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-078-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ. ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3103 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 288...1565
OTHER INFORMATION:
US-08-826-246-3

Query Match 76.0%; Score 15.2; DB 3; Length 3103;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0.

QY 1 gatgactctcctgccttg 20
      ||| ||||| ||||| |||
Db 2031 GGATGCTCTCTGCCCCCTTG 2050

RESULT 15
US-08-944-495-3
; Sequence 3, Application US/08944495
; Patent No. 6087477
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944, 495
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 08/799,910
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 288...1565
; OTHER INFORMATION:
; US-08-944-495-3

```

```

Query Match          76.0%; Score 15.2; DB 3; Length 3103;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

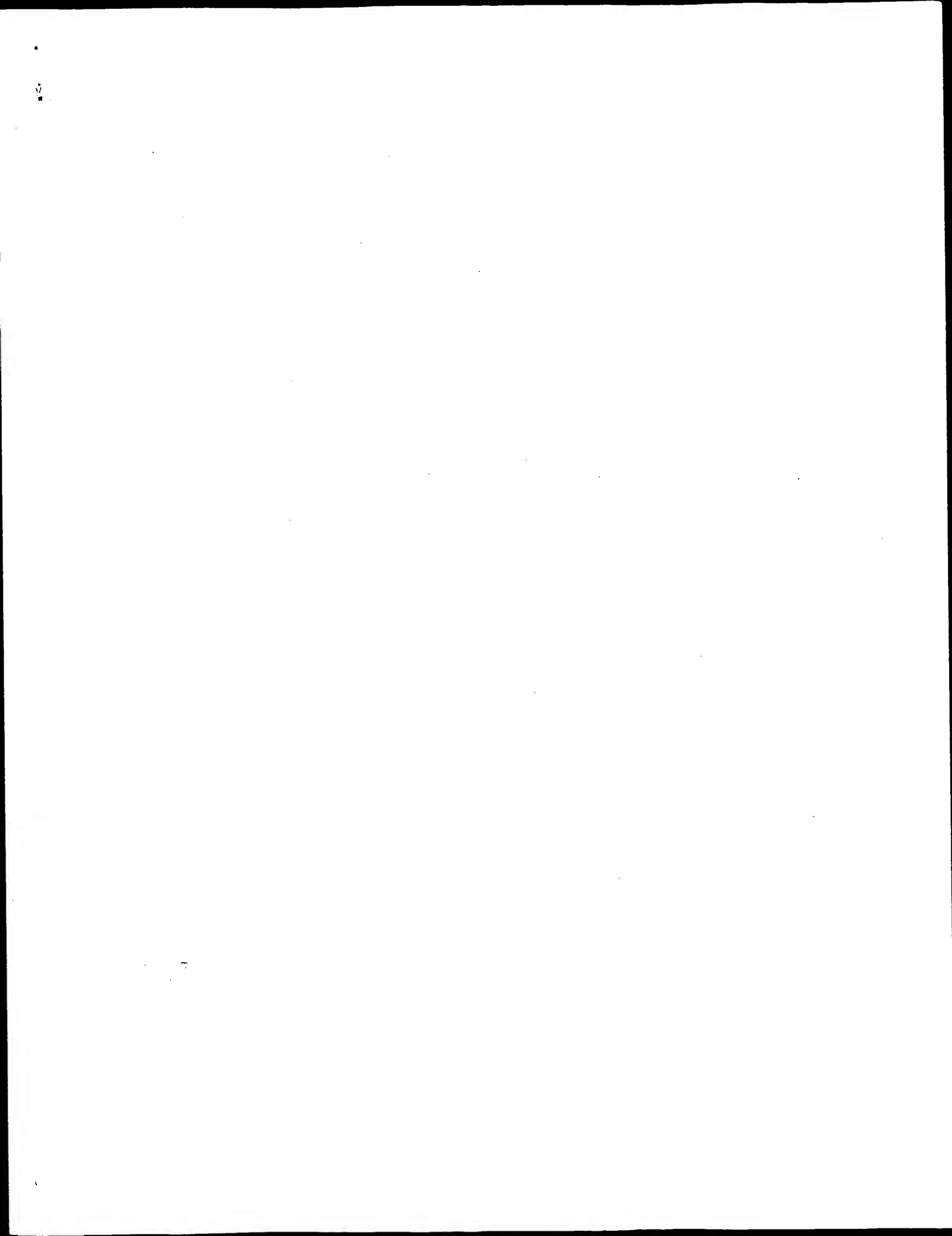
```

```

Oy 1 gatgactcctgaccttg 20
   |||||||||||||
Db 2031 GGTCCTCTCTGCCCCCTTG 2050

```

Search completed: February 24, 2002, 03:01:27
 Job time: 21438 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 03:01:27 ; Search time 212.72 Seconds
(without alignments)
19.164 Million cell updates/sec

Title: US-09-171-553B-14
Perfect score: 18
Sequence: 1 ccacagtcgtacacacag 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/lna/5a.COMB.seq:*
- 2: /cgn2_6/ptodata/2/lna/5b.COMB.seq:*
- 3: /cgn2_6/ptodata/2/lna/5a.COMB.seq:*
- 4: /cgn2_6/ptodata/2/lna/5b.COMB.seq:*
- 5: /cgn2_6/ptodata/2/lna/PCTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18	100.0	3482	3	US-09-111-085-3
2	18	100.0	7333	4	US-08-766-528-2
3	18	100.0	8060	4	US-08-766-528-1
4	18	100.0	8132	4	US-08-766-528-3
5	17	94.4	20	4	US-08-766-528-61
6	17	94.4	28	3	US-09-111-085-15
7	13.8	76.7	1760	3	US-09-150-133-8
8	13.8	76.7	1760	3	US-09-150-141-8
9	13.8	76.7	1760	4	US-09-374-493-8
10	13.8	76.7	1760	4	US-09-374-824-8
11	13.8	76.7	1760	4	US-09-374-824-8
12	13.8	76.7	1760	4	US-09-374-824-8
13	13.8	76.7	1760	4	US-09-374-824-8
14	13.8	76.7	1760	4	US-09-374-824-8
15	13.8	76.7	1760	4	US-09-374-824-8
16	13.8	76.7	1760	4	US-09-374-824-8
17	13.8	76.7	1760	4	US-09-374-824-8
18	13.8	76.7	1760	4	US-09-374-824-8
19	13.8	76.7	1760	4	US-09-374-824-8
20	13.8	76.7	1760	4	US-09-374-824-8
21	13.8	76.7	1760	4	US-09-374-824-8
22	13.8	76.7	1760	4	US-09-374-824-8
23	13.8	76.7	1760	4	US-09-374-824-8
24	13.8	76.7	1760	4	US-09-374-824-8
25	13.8	76.7	1760	4	US-09-374-824-8
26	13.8	76.7	1760	4	US-09-374-824-8
27	13.8	76.7	1760	4	US-09-374-824-8

28	13.4	74.4	1525	2	US-09-006-535-1	Sequence 1, Appl
29	13.4	74.4	3058	4	US-08-818-112-106	Sequence 106, App
30	13.4	74.4	6485	3	US-09-081-320-2	Sequence 2, Appl
31	13.4	74.4	8743	3	US-09-081-320-1	Sequence 1, Appl
32	13.4	74.4	36741	4	US-09-301-665-3	Sequence 3, Appl
33	13.4	74.4	4411529	4	US-09-103-840A-1	Sequence 1, Appl
34	13.2	73.3	187	2	US-08-485-657A-17	Sequence 17, Appl
35	13.2	73.3	187	2	PCT-US95-02303-25	Sequence 25, Appl
36	13.2	73.3	188	2	US-08-485-657A-22	Sequence 22, Appl
37	13.2	73.3	188	5	PCT-US95-02303-21	Sequence 21, Appl
38	13.2	73.3	1621	3	US-09-013-881-14	Sequence 14, Appl
39	13.2	73.3	15378	3	US-08-785-420-1	Sequence 1, Appl
40	13.2	73.3	40138	3	US-09-090-793-12	Sequence 12, Appl
41	13.2	73.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl
42	13	72.2	1342	3	US-08-961-083-181	Sequence 181, App
43	12.8	71.1	78	2	US-08-680-326-46	Sequence 46, Appl
44	12.8	71.1	176	4	US-08-905-223-150	Sequence 150, App
45	12.8	71.1	560	3	US-08-545-809A-62	Sequence 62, Appl

ALIGNMENTS

```

RESULT 1
US-09-111-085-3/c
; Sequence 3, Application US/0911085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match 100.0%, Score 18; DB 3; Length 3482;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccacagtcgtacacacag 18
Db 3439 CCACAGTCGTACACACAG 3422

RESULT 2
US-08-766-528-2/c
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; FILE REFERENCE: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 7333 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-2

Query Match 100.0%; Score 18; DB 4; Length 7333;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
|||||
Db 7279 CCACAGTCGTACACCAG 7262

RESULT 3
US-08-766-528-1/c
Sequence 1, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
City: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match 100.0%; Score 18; DB 4; Length 8060;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
|||||
Db 2546 CCACAGTCGTACACCAG 2529

RESULT 4
US-08-766-528-3/c
Sequence 3, Application US/08766528
Patent No. 6190861

GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
City: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8132 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-3

Query Match 100.0%; Score 18; DB 4; Length 8132;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
|||||
Db 8080 CCACAGTCGTACACCAG 8063

RESULT 5
US-08-766-528-61
Sequence 61, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:

APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-766-528-61

Query Match 94.4%; Score 17; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cacagtcgtacaccacg 18
|||||
Db 1 CACAGTCGTACACCACG 17

RESULT 6
US-09-111-085-15/c
; Sequence 15, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 15
; LENGTH: 28
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: oligonucleotide
US-09-111-085-15

Query Match 94.4%; Score 17; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cacagtcgtacaccacg 18
|||||
Db 28 CACAGTCGTACACCACG 12

RESULT 7
US-09-150-133-8
; Sequence 8, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820,504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 8
; LENGTH: 1760
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-150-133-8

Query Match 76.7%; Score 13.8; DB 3; Length 1760;
Best Local Similarity 88.2%; Pred. No. 1,1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cacagtcgtacaccacg 18
|||||
Db 939 cacagtcgtacaccacg 955

RESULT 8
US-09-150-141-8
; Sequence 8, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820,495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 8
; LENGTH: 1760
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-150-141-8

Query Match 76.7%; Score 13.8; DB 3; Length 1760;
Best Local Similarity 88.2%; Pred. No. 1,1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cacagtcgtacaccacg 18
|||||
Db 939 cacagtcgtacaccacg 955

RESULT 9
US-09-374-493-8
; Sequence 8, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:

APPLICANT: The Board of Regents of the University of Oklahoma
TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
FILE REFERENCE: 5820.546
CURRENT APPLICATION NUMBER: US/09/374,493
EARLIER APPLICATION NUMBER: 09/150,133
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 60/072,994
EARLIER FILING DATE: 1998-01-29
EARLIER APPLICATION NUMBER: PCT/US99/16750
EARLIER FILING DATE: 1999-07-23
NUMBER OF SEQ ID NOS: 45
SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
SEQ ID NO 8
LENGTH: 1760
TYPE: DNA
ORGANISM: Mus musculus
US-09-374-493-8

Query Match 76.7%; Score 13.8; DB 4; Length 1760;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cacagtcgtacacacg 18
||||| | |||||
Db 939 cacagtcgtacacacg 955

RESULT 10
US-09-374-824-8
Sequence 8, Application US/09374824
Patent No. 6207414
GENERAL INFORMATION:
APPLICANT: The Board of Regents of the University of Oklahoma
TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
FILE REFERENCE: 5820.547
CURRENT APPLICATION NUMBER: US/09/374,824
CURRENT FILING DATE: 1999-08-13
EARLIER APPLICATION NUMBER: 09/150,133
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 60/072,994
EARLIER FILING DATE: 1998-01-29
EARLIER APPLICATION NUMBER: PCT/US99/16750
EARLIER FILING DATE: 1999-07-23
NUMBER OF SEQ ID NOS: 45
SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
SEQ ID NO 8
LENGTH: 1760
TYPE: DNA
ORGANISM: Mus musculus
US-09-374-824-8

Query Match 76.7%; Score 13.8; DB 4; Length 1760;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cacagtcgtacacacg 18
||||| | |||||
Db 939 cacagtcgtacacacg 955

RESULT 11
US-09-374-492-8
Sequence 8, Application US/09374492
Patent No. 6207432
GENERAL INFORMATION:
APPLICANT: The Board of Regents of the University of Oklahoma
TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
FILE REFERENCE: 5820.545
CURRENT APPLICATION NUMBER: US/09/374,492
CURRENT FILING DATE: 1999-08-13

EARLIER APPLICATION NUMBER: 09/150,141
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 60/072,994
EARLIER FILING DATE: 1998-01-29
EARLIER APPLICATION NUMBER: PCT/US99/16750
EARLIER FILING DATE: 1999-07-23
NUMBER OF SEQ ID NOS: 45
SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
SEQ ID NO 8
LENGTH: 1760
TYPE: DNA
ORGANISM: Mus musculus
US-09-374-492-8

Query Match 76.7%; Score 13.8; DB 4; Length 1760;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cacagtcgtacacacg 18
||||| | |||||
Db 939 cacagtcgtacacacg 955

RESULT 12
US-08-891-298-2
Sequence 2, Application US/08891298
Patent No. 6300488
GENERAL INFORMATION:
APPLICANT: Gage, Frederick H.
TITLE OF INVENTION: Modified Lepidopteran Receptors
TITLE OF INVENTION: and Hybrid Multi-Functional Proteins for use in Transcript
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/891,298
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2711 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 185...2002
OTHER INFORMATION:
US-08-891-298-2

Query Match 76.7%; Score 13.8; DB 4; Length 2711;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
|||||
DB 1092 CCACAGTCGACACACCAC 1108

RESULT 13

US-08-042-747A-4/C
; Sequence 4, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scintacariello, Franco
; APPLICANT: Hillard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042,747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 269..2941
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..249
; US-08-042-747A-4

Query Match 76.7%; Score 13.8; DB 1; Length 3177;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ccacagtcgtacaccac 18
|||||
DB 166 CGCGGTGTACACCACG 150

RESULT 14

US-08-755-587-27

; Sequence 27, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ. ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5892 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 483..5412
; FEATURE:
; NAME/KEY: exon
; LOCATION: 481..5412
; US-08-755-587-27

Query Match 76.7%; Score 13.8; DB 3; Length 5892;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
|||||
DB 235 CCACAGCATACACCAC 251

RESULT 15

US-08-480-784-20/c
; Sequence 20, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff

```

? APPLICANT: Kamb, Alexander
? APPLICANT: Harshman, Keith D.
? APPLICANT: Shattuck-Eidens, Donna M.
? APPLICANT: Tavligian, Sean V.
? APPLICANT: Wiseman, Roger W.
? APPLICANT: Futreal, P. Andrew
? TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
? TITLE OF INVENTION: Susceptibility Gene
? NUMBER OF SEQUENCES: 85
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
? STREET: 1201 New York Avenue, N.W., Suite 1000
? CITY: Washington
? STATE: DC
? COUNTRY: USA
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/480,784
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/409,305
? FILING DATE: 24-MAR-1995
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/348,824
? FILING DATE: 29-NOV-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/308,104
? FILING DATE: 16-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/300,266
? FILING DATE: 02-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/289,221
? FILING DATE: 12-AUG-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Ihnen, Jeffrey L.
? REGISTRATION NUMBER: 28,957
? REFERENCE/DOCKET NUMBER: 24884-109347
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-962-4810
? TELEFAX: 202-962-8300
? INFORMATION FOR SEQ. ID NO: 20:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 6769 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Homo sapiens
? US-08-480-784-20

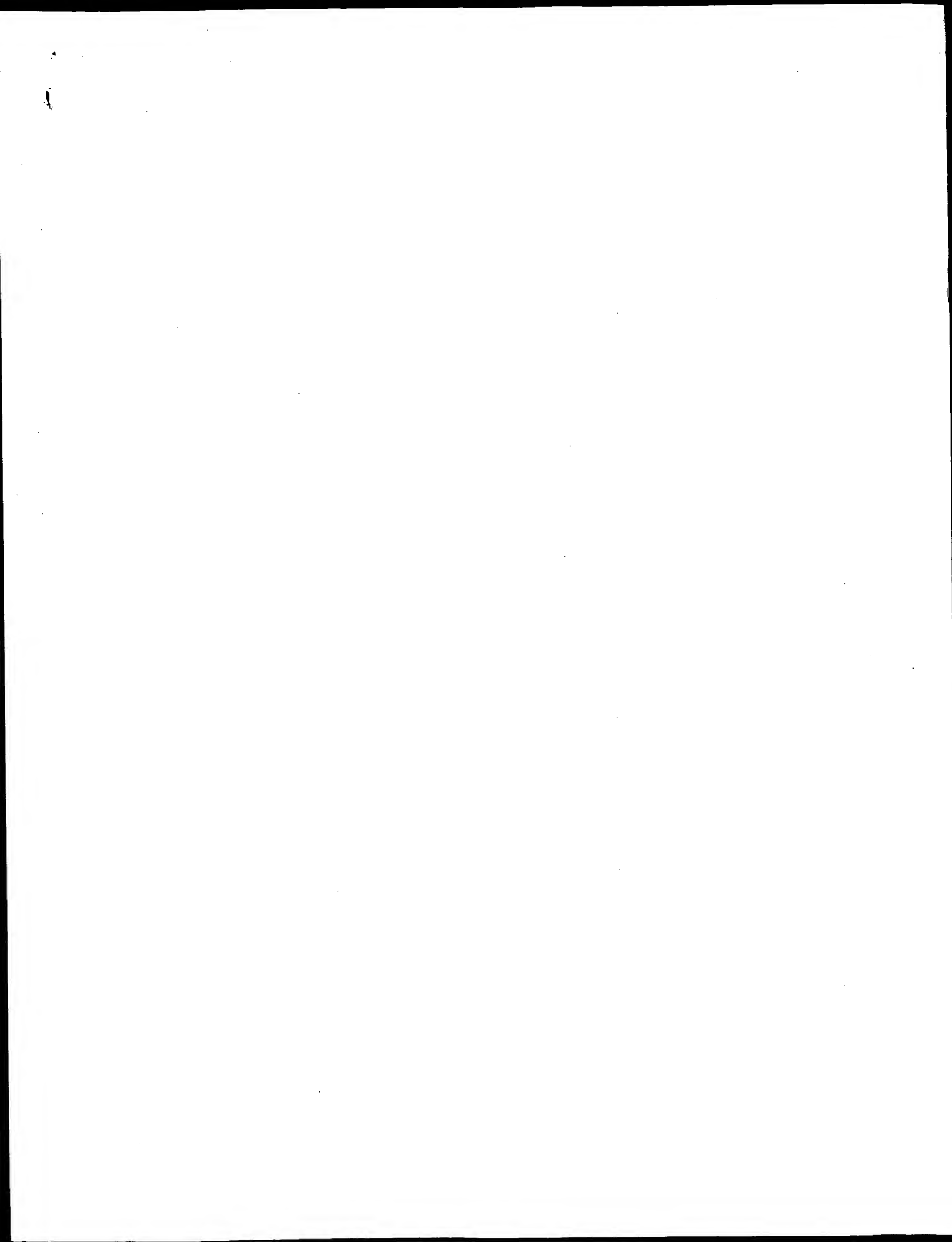
Query Match 76.7%; Score 13.8; DB 1; Length 6769;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccaacatcgtacaccac 17
Db 965 CTACAGCGGTACACCAC 949

```

Mon Feb 25 07:44:10 2002

us-09-171-553b-14.feb22std.rni



FEATURES
Source
POLYA=yes.
Location/Qualifiers
1. .391
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="MI-P-Ay0-nae-c-07-0-UU"
/clone_lib="MI-P-Ay0"
/lab_host="DHI0B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: EcoRI. The MI-P-Ay0 library is derived from placenta. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>. The procedure used to create this library has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=MI-P-Ay0
TAG_TISSUE=placenta
TAG_SEQ=ATTGCG
BASE COUNT 94 a 87 c 71 g 139 t
ORIGIN

Query Match 2.2%: Score 74; DB 11; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e-11;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1984 gtttttagaacaccccttcagagatggtagagagctatcctactaagaagagactca 2043
|||||
Db 250 gtttttagaacaccccttcagagatggtagagagctatcctactaagaagagactca 191
|||||
QY 2044 accgtgtgtgctaa 2057
|||||
Db 190 accgtgtgtgctaa 177
|||||

RESULT 2
BE013835/c 545 bp mRNA EST 09-JUL-2000
LOCUS BE013835
DEFINITION 125326 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE013835
VERSION BE013835.1 GI:8274822
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 545)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68931-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mismatch 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCAGCAGC
Plate: 52 row: B column: 23
Seq primer: ATTAGTGTGACACTATAG.
Location/Qualifiers
1. .545
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"

/tissue_type="pooled"
/lab_host="DHI0B"
/note="Vector: pCMV SPORF6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 128 a 120 c 129 g 168 t
ORIGIN

Query Match 2.2%: Score 74; DB 10; Length 545;
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1984 gtttttagaacaccccttcagagatggtagagagctatcctactaagaagagactca 2043
|||||
Db 216 gtttttagaacaccccttcagagatggtagagagctatcctactaagaagagactca 157
|||||
QY 2044 accgtgtgtgctaa 2057
|||||
Db 156 accgtgtgtgctaa 143
|||||

RESULT 3
B1183551 399 bp mRNA EST 10-JUL-2001
LOCUS B1183551
DEFINITION UNL-P-FN-by-g-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
UNL-P-FN-by-g-02-0-UNL 3', mRNA sequence.
ACCESSION B1183551
VERSION B1183551.1 GI:14657960
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 399)
Caetano,A.R., Johnson,R.K. and Pomp,D.
Generation and sequence characterization of a normalized cDNA
library from swine ovarian follicles
Unpublished (2001)
Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
Oligo-dr track not found. Not I site shown in beginning of sequence
is likely internal to the message.
Seq primer: M13 -29
POLYA=No.

FEATURES
Source
Location/Qualifiers
1. .399
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection
Lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-by-g-02-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DHI0B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI. The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging from 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dr track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
TAG_SEQ=None found"
BASE COUNT 120 a 83 c 92 g 96 t 8 others

ORIGIN

Query Match 1.7%: Score 57: DB 11: Length 399:
 Best Local Similarity 100.0%: Pred. No. 1.4e-06;
 Matches 57: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1920 cgcctactgggaagtgactcactgagtaagccgagtaaacgaacgaacaata 1976
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 116 CGCCTACTGGGAGTGGACTTCACTAGTAGTAAGCCGCGTAATACGGAACAATA 172

RESULT 4
 AM447645 403 bp mRNA EST 25-APR-2001
 LOCUS 89634 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
 DEFINITION
 ACCESSION AM447645
 VERSION AM447645.1 GI:6989432
 KEYWORDS EST.
 SOURCE
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 403)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McCorm,C.G.,
 Perta,G., Holt,L., Karanycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine CDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 v0.960904.e. Vector identified by cross_match with the -minscore 20
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACAGTCACGACG
 Plate: 70 row: L column: 19
 Seq primer: ATTGAGTGACACTATG.
 Location/Qualifiers
 1. 403
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 1BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPOR6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."

BASE COUNT 61 a 134 c 124 g 84 t

ORIGIN

Query Match 1.6%: Score 52: DB 10: Length 403:
 Best Local Similarity 100.0%: Pred. No. 3.5e-05;
 Matches 52: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 2461 gactgagtgagcagcagcgtggaagcagctccggaagagcactactcagag 2512
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 302 GAGTGGTGAGGAGCAGCGCTGAGACAGAGCTCGGAGAGCGCTACTAGAG 353

RESULT 5
 A2891915/c

LOCUS A2891915 451 bp DNA GSS 05-MAR-2001
 DEFINITION RPCI-24-172A20.TV RPCI-24 Mus musculus genomic clone RPCI-24-172A20
 , DNA sequence.
 ACCESSION A2891915
 VERSION A2891915
 KEYWORDS GSS.
 SOURCE
 ORGANISM Mus musculus
 house mouse.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 451)
 Zhao,S., Neriman,W., Malek,J., Shatsman,S., Akınret,B., Levins,M.,
 Tsegaye,G., Geier,K., Krol,M., Shwartsbeyn,A., Gebregorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other_GSSs: RPCI-24-172A20.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdjong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 172 row: A column: 20
 Seq primer: 77
 Class: BAC ends.
 Location/Qualifiers
 1. 451
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-172A20"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 144 a 91 c 106 g 110 t

ORIGIN

Query Match 1.6%: Score 52: DB 13: Length 451:
 Best Local Similarity 100.0%: Pred. No. 3.2e-05;
 Matches 52: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 2420 tgcctgtccacagccttctctctctagctcaagcgcgcagtggtgtgag 2471
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 304 TGCTGCTTCCACAGCCTTGTTCTCTAGGCTCAAGCGCTCGAGTGGGTGAG 253

RESULT 6
 A2380763 549 bp DNA GSS 02-OCT-2000
 LOCUS 1M0136B22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0136B22 R, DNA sequence.
 ACCESSION A2380763
 VERSION A2380763.1 GI:10494463
 KEYWORDS GSS.
 SOURCE
 ORGANISM Mus musculus
 house mouse.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 549)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A., and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 CONTACT: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
 Plate: 0136 row: B column: 22
 Seq primer: CACACAGGAAACAGCATGTGACC
 Class: plasmid ends
 High quality sequence stop: 549.

FEATURES

Source
 Location/Qualifiers
 1..549
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UNGCLM0136B22"
 /clone_1lb="Mouse 10kb plasmid UNGCLM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g1473214|gb|AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

159 a 119 c 119 g 152 t

ORIGIN

Query Match 1.5%; Score 50; DB 13; Length 549;
 Best Local Similarity 100.0%; Pred. No. 9.8e-05;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 228 GTTCCAGGTAACTGAGGCTGGCCAAAGATATTGGGATGATGGA 277

RESULT

7

LOCUS

AZ730616

DEFINITION

RPCT-24-149D1.TJ RPCT-24 Mus musculus genomic clone RPCT-24-149D1,

ACCESSION

AZ730616

VERSION

AZ730616.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus.

REFERENCE

1 (bases 1 to 576)

AUTHORS
 Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akınret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P., and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCT-24
 Unpublished (1999)
 Other_GSSs: RPCT-24-149D1.TJ
 CONTACT: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCT-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orderframe.htm). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 149 row: D column: 1
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Source
 Location/Qualifiers
 1..576
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCT-24-149D1"
 /clone_1lb="RPCT-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI. RPCT-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT

161 a 139 c 121 g 155 t

ORIGIN

Query Match 1.3%; Score 42; DB 13; Length 576;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 2406 acatagctgcatgctgctctccagccttctctctag 2447
 317 ACATACTGCTGATGCTCTTCCACGCTTGTCTCTAG 358

RESULT

8

LOCUS

BH059214

DEFINITION

RPCT-24-326H3.TJ RPCT-24 Mus musculus genomic clone RPCT-24-326H3,

ACCESSION

BH059214

VERSION

BH059214.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus.

REFERENCE

1 (bases 1 to 704)

TITLE

Mouse BAC End Sequences from Library RPCT-24

JOURNAL

Unpublished (1999)

COMMENT

Other_GSSs: RPCT-24-326H3.TJ
 CONTACT: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html

Plate: 326 row: H column: 3
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers

FEATURES

source

1..704
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-326H3"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: PTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT
ORIGIN

224 a 132 c 174 g 174 t

Query Match 1.2% Score 40; DB 13; Length 704;
Best Local Similarity 100.0%; Pred. No. 0.049; 0; Indels 0; Gaps 0;
Matches 40; Conservative 0; Mismatches 0;

Oy 2137 caggtgaagtcaggagctggcgaalatgttgagattgatt 2176
|||||
Db 562 CAGGTAAGTCAGGACTGGCCAGATATTGGGCAATTGATT 601

RESULT 9
LOCUS BF442561 100 bp mRNA EST 01-DEC-2000
DEFINITION 259350 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF442561
VERSION BF442561.1 GI:11502653
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 100)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.

REFERENCE

AUTHORS Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

JOURNAL

COMMENT Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smt@meaill.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.

FEATURES

source

1..100
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"

/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT
ORIGIN

29 a 24 c 26 g 21 t

Query Match 1.2% Score 39; DB 11; Length 100;
Best Local Similarity 100.0%; Pred. No. 0.46; 0; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 0;

Oy 1348 ggggcatcattaacaagggtgttcttaaccca 1386
|||||
Db 55 GGGGCCATCTATTAACAAAGGGGTGTCTTACCTCAGCA 93

RESULT 10
LOCUS AO920035 482 bp DNA GSS 21-DEC-1999
DEFINITION RPCI-23-273D7.TV RPCI-23 Mus musculus genomic clone RPCI-23-273D7, DNA sequence.
ACCESSION AO920035
VERSION AO920035.1 GI:6609038
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 482)
Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatman,S., Akiret,
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.

REFERENCE

AUTHORS Mouse BAC End Sequences from Library RPCI-23

JOURNAL

COMMENT Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 273 row: D column: 7
Seq primer: T7
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..482
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-273D7"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT
ORIGIN

128 a 91 c 106 g 154 t 3 others

Query Match

1.1% Score 37; DB 13; Length 482;

Best Local Similarity 100.0%; Pred. NO. 0.46;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2420 tgcgtcttccacgcttgcctctagctcaagc 2456
 Db 29 TGCCTCTTCCACGCTTGTCTCTAGCTCAAGC 65

RESULT 11

AA529580

545 bp mRNA

EST

22-JUL-1997

LOCUS

v41404.r1 Beddington mouse embryonic region Mus musculus cDNA

DEFINITION

clone IMAGE:906343 5' similar to SW:POL_GALV P21414 POL POLYPEPTIDE

; mRNA sequence.

ACCESSION

AA529580

AA529580.1

GI:2272286

VERSION

EST.

KEYWORDS

house mouse,

ORGANISM

Mus musculus

REFERENCE

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 545)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE

The WashU-HM Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

Contact: Maria M/Mouse EST Project
 WashU-HM Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LML; contact the
 IMAGE Consortium (info@image.lml.gov) for further information.
 MGI:527007
 Seq primer: -28m3 rev1 ET from Amersham
 High quality sequence stop: 326.

FEATURES

Location/Qualifiers

source

1..545

/organism="Mus musculus"

/strain="C57BL/6 x DBA"

/db_xref="taxon:10090"

/clone="IMAGE:906343"

/clone_1ib="Beddington mouse embryonic region"

/sex="pooled"

/tissue_type="embryo"

/dev_stage="7.5dpc"

/lab_host="DH12S"

/note="organ: whole embryo; Vector: pCMV-SPORT; Site:1;
 Salt: Site:2: NotI; Cloned unidirectionally. Primer:
 Oligo dt. Gastrulating embryos were collected at 7.5dpc
 from C57BL/6 x DBA matings, excluding embryos that had
 developed head folds and all extremembryonic tissues.
 Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
 Referenced in Development 121, 2479-2489 (1995)."

BASE COUNT

143 a

130 c

127 g

145 t

ORIGIN

Query Match 1.1%; Score 36; DB 10; Length 545;
 Best Local Similarity 100.0%; Pred. NO. 0.78;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2420 tgcgtcttccacgcttgcctctagctcaagc 2455
 Db 176 TGCCTCTTCCACGCTTGTCTCTAGCTCAAGC 211

RESULT 12

A2479717/c
 LOCUS A2479717 575 bp DNA
 DEFINITION 1M0300M05R mouse 10kb plasmid UUC1M library Mus musculus genomic
 clone UUC1M0300M05 R, DNA sequence.
 A2479717
 VERSION A2479717.1
 KEYWORDS GI:10639886
 SOURCE house mouse.
 ORGANISM house mouse.

REFERENCE

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 575)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duvall, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0300 row: M column: 05
 Seq primer: CACACAGGAAACAGCTAGAAC
 Class: plasmid ends
 High quality sequence stop: 575.

FEATURES

Location/Qualifiers

source

1..575

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUC1M0300M05"

/clone_1ib="Mouse 10kb plasmid UUC1M library"

/sex="Male"

/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"

/note="Vector: pMD42hv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g11473211[9b]AT129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT

136 a

140 c

137 g

162 t

ORIGIN

Query Match 1.1%; Score 36; DB 13; Length 575;
 Best Local Similarity 100.0%; Pred. NO. 0.75;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2203 caaagctcagacagcagtagagatgaagaac 2238
 Db 160 CAAAGCTCAGGACAGCTAGAGAGATGAAGAACC 125

RESULT 13
 A2415280/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 SOURCE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 source

A2415280 583 bp DNA
 1M0190M06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0190M06 F, DNA sequence.
 A2415280
 A2415280.1 GI:10539293
 GSS.
 Mus musculus
 house mouse.
 Mus musculus
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 583)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0190 row: M column: 06
 Seq primer: CGTTGTAACGACGCGCCGT
 Class: plasmid ends
 High quality sequence stop: 583.
 Location/Qualifiers
 1..583
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0190M06"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g11473211419b1AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 146 a 149 c 130 g 158 t
 ORIGIN

Query Match 1.1%; Score 36; DB 13; Length 583;
 Best Local Similarity 100.0%; Pred. No. 0.74;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2203 caaagctcagacagtgatgagatgaataaac 2238
 ||||||||||||||||||||||||||||||||||||
 Db 193 CAAAGCTCAGACAGGTAGAGAGATGAATAGAAC 158

RESULT 14
 A2944719
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 source

A2944719 599 bp DNA
 2M0205L23R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0205L23 R, DNA sequence.
 A2944719
 A2944719.1 GI:13809008
 GSS.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 599)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0205 row: L column: 23
 Seq primer: CACACAGCAACAGCTAGAAC
 Class: plasmid ends
 High quality sequence stop: 599.
 Location/Qualifiers
 1..599
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0205L23"
 /clone_lib="Mouse 10kb plasmid UUGC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g11473211419b1AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 160 a 136 c 151 g 152 t
 ORIGIN

Query Match 1.1%; Score 36; DB 13; Length 599;
 Best Local Similarity 100.0%; Pred. No. 0.72;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2203 caaagctcagacagtgatgagatgaataaac 2238
 ||||||||||||||||||||||||||||||||||||
 Db 334 CAAAGCTCAGACAGGTAGAGAGATGAATAGAAC 369

RESULT 15

AZ859488

LOCUS

610 bp DNA

GSS 21-FEB-2001

2M0165D07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION

AZ859488

VERSION

AZ859488.1 GI:13053858

KEYWORDS

GSS.

SOURCE

house mouse

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE

1 (bases 1 to 610)

JOURNAL

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,

COMMENT

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

JOURNAL

M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

COMMENT

and Wright,D., Weiss,R.

JOURNAL

Mouse whole genome scaffolding with paired end reads from 10kb

COMMENT

plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss

JOURNAL

University of Utah Genome Center

COMMENT

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

JOURNAL

84112, USA

COMMENT

Tel: 801 585 5606

JOURNAL

Fax: 801 585 7177

COMMENT

Email: ddunn@genetics.utah.edu

JOURNAL

Insert length: 10000 Std Error: 0.00

COMMENT

Plate: 0165 row: D column: 07

JOURNAL

Seq primer: CGTTGTAACGACGCGCAGT

COMMENT

Class: Plasmid ends

JOURNAL

High quality sequence stop: 610.

COMMENT

Location/Qualifiers

JOURNAL

1. 610

COMMENT

/organism="Mus musculus"

JOURNAL

/strain="C57BL/6J"

COMMENT

/db_xref="taxon:10090"

JOURNAL

/clone="UUGC2M0165D07"

COMMENT

/clone_lib="Mouse 10kb plasmid UUGC1M library"

JOURNAL

/sex="Male"

COMMENT

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

BASE COUNT

194 a 145 c 130 g 141 t

ORIGIN

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match

1.1%; Score 36; DB 13; Length 610;

Best Local Similarity

100.0%; Pred. No. 0.71;

Matches

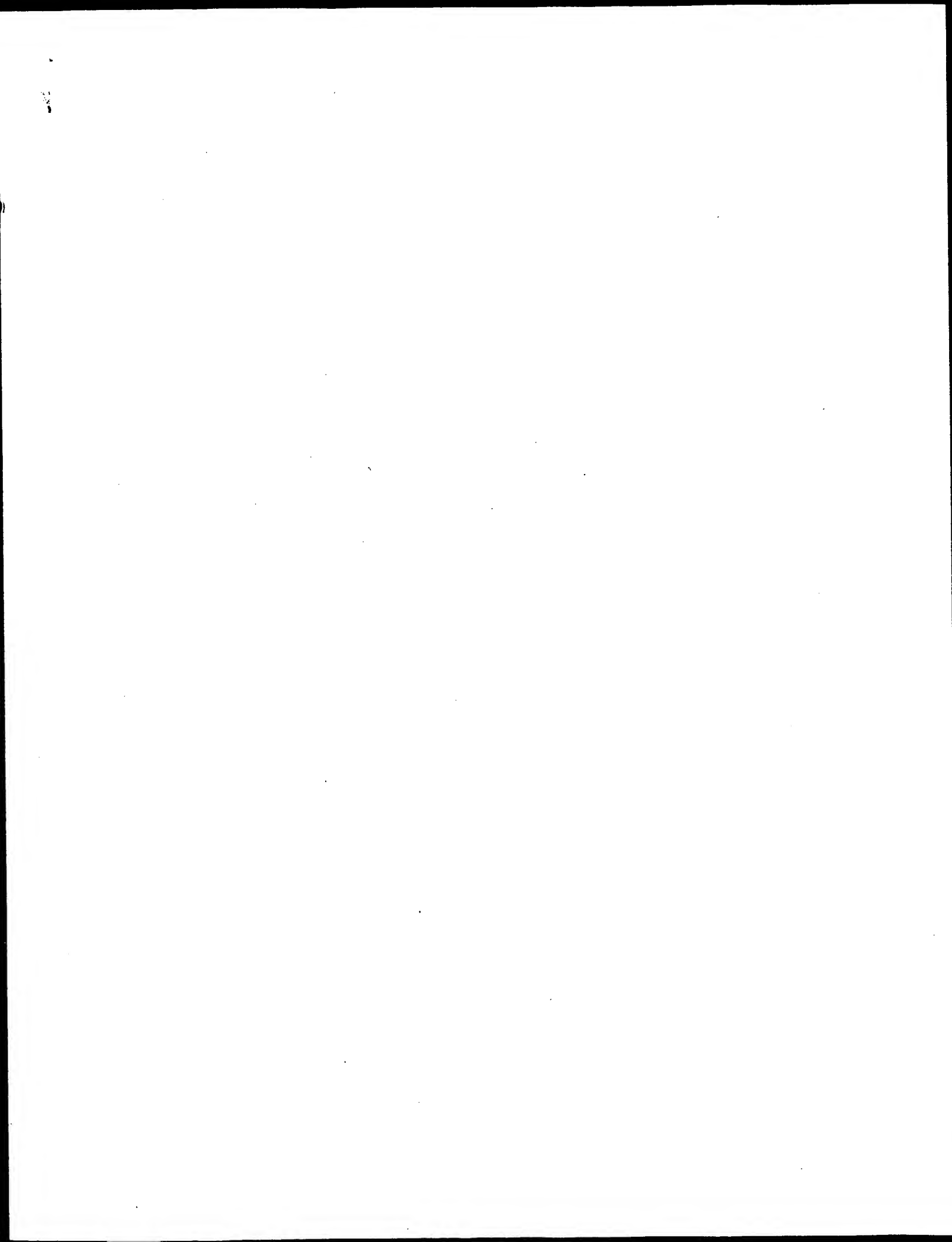
36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query

2420 tgcgtctccagccttctctaggtcaagg 2455

DB 410 TGCCTCTTCCAGCCTTTGTTCTCTAGGCTCAAG 445

Search completed: February 24, 2002, 04:28:16
Job Time: 24042 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 04:28:16 ; Search time 6170.56 Seconds
(without alignments)
14295.639 Million cell updates/sec

Title: US-09-171-553b-3

Perfect score: 8209

Sequence: 1 gtggtgctacgactggtggcc.....aaaaaaaaaaaaaaaaaaaa 8209

Scoring table:

OLIGO.NUC
Gapop 60.0 , Gapept 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	190	2.3	564	11	BI182930 UNL-P-FN-
2	140	1.7	466	11	BF441468 257862 MA
3	140	1.7	467	11	BF441465 257856 MA
C 4	130	1.6	532	11	BF709893 MI-P-A11-
5	126	1.5	408	11	BF711018 MI-P-A11-
6	102	1.2	564	11	BI181019 UNL-P-FN-
7	102	1.2	564	11	BI181029 UNL-P-FN-
8	96	1.1	376	11	BF441469 257863 MA
9	92	1.1	558	11	BI360879 388957 MA
C 10	92	1.1	602	11	BF712162 MI-P-B6-a
11	80	1.0	452	11	BF441466 257857 MA
12	78	1.0	549	10	AW657531 110922 MA

13	77	0.9	857	11	BI184013 UNL-P-FN-
C 14	74	0.9	391	11	BF709087 MI-P-AV0-
C 15	74	0.9	545	10	BE013835 125326 MA
C 16	72	0.9	229	11	BF443400 260872 MA
C 17	69	0.8	468	11	BI186066 UNL-P-FN-
C 18	69	0.8	499	11	BF704354 MI-P-O3-a
C 19	69	0.8	521	11	BI399852 MI-P-AV1-
C 20	69	0.8	730	11	BI185535 UNL-P-FN-
C 21	68	0.8	468	11	BI182727 UNL-P-FN-
C 22	64	0.8	447	11	BI401063 MI-P-CP0-
C 23	64	0.8	485	11	BF703671 MI-P-B4-a
C 24	64	0.8	554	11	BF703805 MI-P-O2-a
C 25	63	0.8	575	11	BI183356 UNL-P-FN-
C 26	59	0.7	465	11	BI399234 MI-P-AV1-
C 27	57	0.7	399	11	BI183551 UNL-P-FN-
C 28	55	0.7	481	10	AM231947 2C22N1 Su
C 29	52	0.6	403	10	AM447645 89634 MAR
C 30	52	0.6	451	13	A2891915 RPCI-24-1
C 31	51	0.6	545	11	BF701947 MI-P-E3-a
C 32	50	0.6	494	11	BI182742 UNL-P-FN-
C 33	49	0.6	549	11	A2380763 1M0136B22
C 34	49	0.6	300	11	BI184279 UNL-P-FN-
C 35	48	0.6	345	10	AM358862 43799 MAR
C 36	48	0.6	595	11	BI185465 UNL-P-FN-
C 37	47	0.6	294	11	BF713588 MI-P-O2-a
C 38	47	0.6	579	11	BI183723 UNL-P-FN-
C 39	45	0.5	286	13	A2771375 1M0573C02
C 40	45	0.5	575	13	BH025657 RPCI-24-3
C 41	45	0.5	599	13	A2041781 RPCI-23-2
C 42	44	0.5	173	10	AM435835 75059 MAR
C 43	44	0.5	341	10	AM416859 52601 MAR
C 44	42	0.5	576	13	A2730616 RPCI-24-1
C 45	41	0.5	256	10	AM308385 3994 MARC

ALIGNMENTS

RESULT 1
BI182930 564 bp mRNA EST 10-JUL-2001
LOCUS UNL-P-FN-br-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
DEFINITION UNL-P-FN-br-h-08-0-UNL 3', mRNA sequence.
BI182930
ACCESSION BI182930.1 GI:14657339
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Caetano, A.R., Johnson, R.K. and Pomp, D.
Generation and sequence characterization of a normalized cDNA
library from swine ovarian follicles
Unpublished (2001)
Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a donatide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized porcine ovarian follicles library
Seq primer: M13 -29
POLY-A-Yes

FEATURES
source
Location/Qualifiers
1..564
/organism="Sus scrofa"

TEL: 402 762 4300
Fax: 402 762 4390

FEATURES Location/Qualifiers

source 1. .467
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_1ib="MARC 2P1G"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
 library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."
 BASE COUNT 115 a 118 c 127 g 107 t
 ORIGIN

Query Match 1.7%; Score 140; DB 11; Length 467;
 Best Local Similarity 100.0%; Pred No. 4.3e-26;
 Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1359 caagctctgacagacactcttcaacacggagagagagaaatctgttgaagagctaga 1418
 |||||||
 Db 61 CAGCTGCTGCAGACACTCTTCACACGAGAGAGAGAGATTCTGTTAGAGCGCTAGA 120

QY 1419 aaaaatgtctctggggcgagcgagcgacccagcagttgcaaaatgagattgacatggga 1478
 |||||||
 Db 121 AAAAATGTTCTCTGGGGCGAGCGGCGACCCAGTCGAAATGAGATTGACATGGGA 180

QY 1479 ttccccctgactgccccgg 1498
 |||||||
 Db 181 TTTCCTTGACTGCCCCGG 200

RESULT 4
 BE709993 532 bp mRNA EST 02-JAN-2001
 LOCUS BE709993/c
 DEFINITION MI-P-A11-nga-b-06-0-UI.s1 MI-P-A11 Sus scrofa cDNA clone
 BE709993
 VERSION BE709993.1 GI:12009470
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 532)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tugale CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kilgus Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ktugale@iastate.edu
 Oligo-dT track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA library Preparation: M.B.
 Soares Lab, University of Iowa. EST sequencing: M.B. Soares Lab,
 University of Iowa. Clone distribution: clones will be available
 through Research Genetics (www.resgen.com) The following repetitive
 elements were found in this cDNA sequence: 59-132,
 >GC-rich#low-complexity
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1. .532
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-A11-nga-b-06-0-UI"
 /clone_1ib="MI-P-A11"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7R3D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: EcoRI; The MI-P-A11
 library is normalized library derived from the MI-P-A10
 library, ultimately derived from placenta tissue. For a
 detailed description of the library from which this clone
 was derived, please visit our web site at
 http://pigest.genome.iastate.edu/. The procedure used to
 create this library has been previously described (Bonaldo
 , Lennon and Soares, Genome Research 6: 791-806, 1996)
 TAG_SEQ=None found"
 BASE COUNT 124 a 184 c 154 g 70 t
 ORIGIN

Query Match 1.6%; Score 130; DB 11; Length 532;
 Best Local Similarity 99.4%; Pred No. 1.3e-23;
 Matches 180; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 364 cccctgggagacgccccggaggtgggagagcagagacgctgtgtctctactgtc 423
 |||||||
 Db 427 CCCTGGGGAGACCCCGGGAGGTGGGAGAGCCAGGAGCGCTGTGTCTCTGTGTC 368

QY 424 ggtcagagagacccgaattctgtgtgagcgaaagcttccccctcgcgccgtccgact 483
 |||||||
 Db 367 GGTCAAGAGACCGACGCTCTGTGTTGAAGCGAAAGCTTCCCTCCGGCGCGTCCACT 308

QY 484 ctttgacctgtgtgagagacgagcgagcggtgtgtctgtgactgtgttctg 543
 |||||||
 Db 307 CTTTTCCTCTGCTTGTGGAAGACGCGGAGCGGCTGTGTGTGATGCTGTGTTCTG 248

QY 544 t 544
 Db 247 T 247

RESULT 5
 BE711018 408 bp mRNA EST 02-JAN-2001
 LOCUS BE711018/c
 DEFINITION MI-P-A11-nrb-h-02-0-UI.s1 MI-P-A11 Sus scrofa cDNA clone
 BE711018
 VERSION BE711018.1 GI:12010495
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 408)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tugale CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kilgus Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ktugale@iastate.edu
 Oligo-dT track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA library Preparation: M.B.
 Soares Lab, University of Iowa. EST sequencing: M.B. Soares Lab,
 University of Iowa. Clone distribution: clones will be available
 through Research Genetics (www.resgen.com) The following repetitive
 elements were found in this cDNA sequence: 59-132,
 >GC-rich#low-complexity
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1. .408
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"

```

/clone="MI-P-AV1-ntd-h-02-0-UT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: EcoRI. The MI-P-AV1
library is normalized library derived from the MI-P-AV0
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
http://pigst.genome.iastate.edu/. The procedure used to
create this library has been previously described (Bonaldo
, Lennon and Soares, Genome Research 6: 791-806, 1996)
TAG_SEQ=None found"
BASE COUNT      94 a      139 c      122 g      52 t      1 others
ORIGIN
Query Match      1.5%; Score 126; DB 11; Length 408;
Best Local Similarity 100.0%; Pred. No. 1.6e-22;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 419 ctctgcgtcaagagaccagctctgttgaagcgaagctcccccgcgcgcgtc 478
|||||
DB 372 CTGTCGTGACAGAGACCGAGTCTGTGTTGAAGCGAAGCTTCCCTCCGCGCCGCTC 313
|||||
QY 479 cgaactcttgcctgttgygaagacgcgcgcgcgcgtgtgtctgatactgttgc 538
|||||
DB 312 CGACTCTTTGCTGCTGTGAGACGCGACGCGGCTGCTGCTGATCTGTGTGT 253
|||||
QY 539 ttctgt 544
|||||
DB 252 TTCTGT 247

RESULT 6
BI181099
LOCUS      BI181099      564 bp      mRNA      EST      10-JUL-2001
DEFINITION UNL-P-FN-ab-d-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
ACCESSION  BI181099
VERSION     BI181099
KEYWORDS   EST.
SOURCE     GI:14655508
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 564)
            Caetano, A.R., Johnson, R.K. and Pomp, D.
            Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
            Unpublished (2001)
JOURNAL    Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found. Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA-No.

FEATURES
    source
        1..564
        /organism="Sus scrofa"
        /strain="University of Nebraska, Lincoln Swine Selection
        Lines"
        /db_xref="taxon:9823"
        /clone="UNL-P-FN-ab-d-02-0-UNL"
        /clone_lib="UNL-P-FN"
        /dev_stage="ADULT"
        /lab_host="DH10B (Life Technologies)"
        /note="Vector: pT73D-Pac (Pharmacia) with a modified
        polylinker. Site 1: Not I; Site 2: Eco RI. The UNL-P-FN
        library is a normalized library representing porcine
        ovarian follicles, ranging between 2.0 to 10.0 mm in
        diameter, collected during 7 days of the follicular phase
        of the pig estrous cycle. This library was derived from
        the library UNL-P-F2. The tag is a string of 5-6
        nucleotides present between the Not I site and the
        oligo-dT track. The library was constructed as described
        by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
        , 1996.
        TAG_SEQ=None found"

```

```

polylinker. Site 1: Not I; Site 2: Eco RI. The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F2. The tag is a string of 5-6
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996.
TAG_SEQ=None found"
BASE COUNT      112 a      196 c      143 g      112 t      1 others
ORIGIN
Query Match      1.2%; Score 102; DB 11; Length 564;
Best Local Similarity 100.0%; Pred. No. 1.4e-16;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1204 gccaatggagcccccagatattggcccttctctctgcagatcctataatggaaa 1263
|||||
DB 372 GCCAATTGACAGCCCCCTCGATATGGCCCTTTCTCTGCAGATCTCTATATTGGA 431
|||||
QY 1264 ctaaccatcccttctcggagatcccaacgcctcagg 1305
|||||
DB 432 CTAAACCATCCCTTTCTCGAGGATCCCAACGCTCAGG 473
|||||

RESULT 7
BI186129
LOCUS      BI186129      564 bp      mRNA      EST      10-JUL-2001
DEFINITION UNL-P-FN-ab-d-02-0-UNL.s2 UNL-P-FN Sus scrofa cDNA clone
ACCESSION  BI186129
VERSION     BI186129
KEYWORDS   EST.
SOURCE     GI:14660538
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 564)
            Caetano, A.R., Johnson, R.K. and Pomp, D.
            Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
            Unpublished (2001)
JOURNAL    Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found. Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA-No.

FEATURES
    source
        1..564
        /organism="Sus scrofa"
        /strain="University of Nebraska, Lincoln Swine Selection
        Lines"
        /db_xref="taxon:9823"
        /clone="UNL-P-FN-ab-d-02-0-UNL"
        /clone_lib="UNL-P-FN"
        /dev_stage="ADULT"
        /lab_host="DH10B (Life Technologies)"
        /note="Vector: pT73D-Pac (Pharmacia) with a modified
        polylinker. Site 1: Not I; Site 2: Eco RI. The UNL-P-FN
        library is a normalized library representing porcine
        ovarian follicles, ranging between 2.0 to 10.0 mm in
        diameter, collected during 7 days of the follicular phase
        of the pig estrous cycle. This library was derived from
        the library UNL-P-F2. The tag is a string of 5-6

```

nucleotides present between the Not I site and the oligo-dr track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.

BASE COUNT 112 a 196 c 144 g 112 t
TAG_SEQ=None found*

Query Match 1.2%; Score 102; DB 11; Length 564;
Best Local Similarity 100.0%; Pred. No. 1.4e-16;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 gccaatgagagcccccagatgagccttctctcgcagatccttaattggaana 1263
|||||
Db 372 GCCAATGGACGCCCTTCACATATTGGCCCTTTCTCTCAGATCTAATAATGGAAAA 431

QY 1264 ctaccatccccctctctcgaagatcccaacgctcaag 1305
|||||
Db 432 CTACCATCCCTTCTCTCGAGAGATCCCAAGCCTCAGG 473

RESULT 8
LOCUS BF441469 376 bp mRNA EST 01-DEC-2000
DEFINITION 257663 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF441469
VERSION BF441469.1 GI:11501633
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 376)
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keeler, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 86 row: D column: 9
Seq primer: ATTTAGTGACACTATAG.
Location/Qualifiers

FEATURES
source 1..376

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 91 a 94 c 108 g 83 t

ORIGIN

Query Match 1.2%; Score 96; DB 11; Length 376;
Best Local Similarity 100.0%; Pred. No. 6.4e-15;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1359 cagctgctgcagacactctcaacacgagagagagaattcgttagagagctaga 1418

|||||
Db 61 CAGCTGTCGACAGACCTCTTCACAAACGAGAGAGAGAAATCTGTAGAGCTAGA 120

QY 1419 aaaatgttcttgaggccgagcagcgagaccacgcag 1454
|||||
Db 121 AAAAATGTTCCGTGGGGCGGAGCGGCGACCGACGCGAG 156

RESULT 9
LOCUS B1360879 558 bp mRNA EST 01-AUG-2001
DEFINITION 388957 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION B1360879
VERSION B1360879.1 GI:15056907
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 558)
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keeler, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 142 row: B column: 4
Seq primer: ATTTAGTGACACTATAG.
Location/Qualifiers

FEATURES
source 1..558

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 129 a 123 c 119 g 187 t

ORIGIN

Query Match 1.1%; Score 92; DB 11; Length 558;
Best Local Similarity 100.0%; Pred. No. 4.6e-14;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8098 gcaatggagggcagagctctaccacccgcggtgtgtacgactgtggagccagcgact 8157
|||||

Db 132 GCACCTGGGGCGGCGAGTCTCTACCCCTGCGGTGTACGACTGTGGCGCGCGCT 191

QY 8158 tggatataaatccctctgtctgtttgcatcaa 8189
|||||

Db 192 TGGATATAAAATCTCTTCTGCTTTGCAATCAA 223

RESULT 10
LOCUS BF712162 602 bp mRNA EST 02-JAN-2001
DEFINITION MI-P-E6-acx-e-11-1-UM.s1 MI-P-E6 Sus scrofa cDNA clone
ACCESSION MI-P-E6-acx-e-11-1-UM 3', mRNA sequence.
BF712162

VERSION BF712162.1 GI:12011595
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 602)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tugle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ktugle@iastate.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 non-normalized fetus at gestational day 20 library cDNA library
 Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
 Research Center, Department of Animal Science, University of
 Missouri-Columbia, 65211 Clone distribution: clones will be
 available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLY=A=yes.

FEATURES
 source
 Location/Qualifiers
 1..602
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="M1-P-E6-acc-e-11-1-UM"
 /clone_1ib="M1-P-E6"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: EcoRI; The M1-P-E6
 library is derived from fetus at gestational day 20. For
 a detailed description of the library from which this
 clone was derived, please visit our web site at
 http://pigest.genome.iastate.edu/
 TAG_LIB=M1-P-E6
 TAG_TISSUE=fetus at gestational day 20
 TAG_SEQ=AGCAA"

BASE COUNT 141 a 146 c 152 g 162 t 1 others
 ORIGIN

Query Match 1.1%; Score 92; DB 11; Length 602;
 Best Local Similarity 99.3%; Pred. No. 4.3e-14;
 Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8060 ggccttggtgaaccataaagctgtccgattccgactcggggccgactcct 8119
 DB 143 GGCCTTGTTGTGAACCCATAAAGCTGCCGATTCGCCACGCGGCGCGAGTCCTCT 84
 QY 8120 accctggcgtgtgtacgtatgggccccagcgagcgttggaataaactctctgcgtg 8179
 DB 83 ACCCGTCGGCGGTACGACTGTGGGCCCCAGCGCGCTTGGAATAAATCTCTTGTG 24
 QY 8180 ttgtcatcaaaaaaaaaaaaaa 8202
 DB 23 TTTCATCAAAAAAAAAAAAAA 1

RESULT 11
 BF441466 452 bp MRNA EST 01-DEC-2000
 LOCUS
 DEFINITION 257857 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.

ACCESSION BF441466
 VERSION BF441466.1 GI:11501630
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 452)
 AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
 and Keefe,J.W.
 TITLE Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 JOURNAL Unpublished (2000)
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402/762 4366
 Fax: 402/762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross-match with the -minscore 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCACGACG
 Plate: 86 row: C column: 9
 Seq primer: ATTTCGTCACATATAG.

FEATURES
 source
 Location/Qualifiers
 1..452
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_1ib="MARC 2P1G"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 107 a 118 c 125 g 102 t
 ORIGIN

Query Match 1.0%; Score 80; DB 11; Length 452;
 Best Local Similarity 100.0%; Pred. No. 5.8e-11;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1419 aaaatgttcctggggccgacggcgaccccgagtgcaaaatgagatgacatgga 1478
 DB 120 AAAAATGTTCTGGGGCGACGGCGACCCACGAGTTGCAAAATGAGATTGACATGGGA 179
 QY 1479 ttcccttgactgcgcccg 1498
 DB 180 TTTCCTTGACTGCGCCCG 199

RESULT 12
 AM657531 549 bp MRNA EST 25-APR-2001
 LOCUS
 DEFINITION 110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AM657531
 VERSION AM657531.1 GI:7423429
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 549)
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G.,
 Pettea,G., Holt,I., Karanymcheva,S., Liang,F., Quackenbush,J. and
 Keefe,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)

MEDLINE 21180013

COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 94 row: E column: 12
Seq primer: ATTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..549
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1B0Y"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 145 a 144 c 121 g 139 t

ORIGIN

Query Match 1.0%; Score 78; DB 10; Length 549;
Best Local Similarity 99.2%; Pred. No. 1.6e-10;
Matches 128; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6640 aaatcaatgtgacaaagagcatagaatcaatgacatgaggggtccggaataagcgt 6699
|||||
Db 159 AAATCATATGTGACCAAGAGCATAGAAATCAATGTACATGGGGTCCCGAATAAGCTTC 218
|||||

QY 6700 accctacatgaatttcgggaagggacatgacataggaagaactcccatccaccacaa 6759
|||||
Db 219 ACCCTCATCTGAAGTTTCCGGGAAGGACATGATAGGAAAGCTCCCATCCACACAA 278
|||||

QY 6760 cacccttgc 6768
|||||
Db 279 CACCTTTGC 287
|||||

RESULT 13
LOCUS B1184013 857 bp mRNA EST 10-JUL-2001
DEFINITION UNL-P-FN-cf-b-12-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
ACCESSION B1184013
VERSION B1184013
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 857)
Caetano, A.R., Johnson, R.K. and Pomp, D.
Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
Unpublished (2001)
Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu

Oligo-dt track not found. Not 1 site shown in beginning of sequence is likely internal to the message.
Seq primer: M13 -29
POLYA-NO.

FEATURES
Location/Qualifiers
1..857
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection Lines"
/db_xref="taxon:9823"
/clone_lib="UNL-P-FN-cf-b-12-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT713D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
TAG_SEQ=None found"

BASE COUNT 152 a 217 c 248 g 236 t 4 others

ORIGIN

Query Match 0.9%; Score 77; DB 11; Length 857;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 gcttgaataaatactctctgttgcataagaccgtctcgtgagtgattgg 87
|||||
Db 38 GCTTGAATATAAATCTCTGCTGTTGCATCAAGACCGTTTCGTGAGTATTGGG 97
|||||

QY 88 gttgcgcctctcgtgag 104
|||||
Db 98 GTTGCCTCTTCCGAG 114
|||||

RESULT 14
LOCUS BF709087/c 391 bp mRNA EST 02-JAN-2001
DEFINITION MI-P-AV0-nae-c-07-0-UI.s1 MI-P-AV0 Sus scrofa cDNA clone
ACCESSION BF709087
VERSION BF709087
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 391)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Tugale CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152944201
Email: cktugale@iastate.edu
The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the

non-normalized placenta library cDNA library Preparation: M.B. Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab, University of Iowa Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers
1..391
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="MI-P-Ay0-nae-c-07-0-01"
/lab_host="MI-P-Ay0"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: EcoRI; The MI-P-Ay0 library is derived from placenta. For a detailed description of the library from which this clone was derived, please visit our web site at <http://p1gest.genome.lastate.edu/>. The procedure used to create this library has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=MI-P-Ay0
TAG_SEQ=ATTGG
BASE COUNT 94 a 87 c 71 g 139 t
ORIGIN

Query Match 0.9%; Score 74; DB 11; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4935 gttttgtagacaccttcagagatggtagagcttatcctactaagaagacttca 4994
|||||
DB 250 GTTTTGTAGACACCTTTTCAGATGGGTAGAGCTTATCCTACTAAGAAGACTTCA 191
|||||
QY 4995 accgtgtggtctaa 5008
|||||
DB 190 ACCGTGTGCTTAA 177
|||||

RESULT 15
BE013835/c 545 bp mRNA EST 09-JUL-2000
LOCUS
DEFINITION 125326 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE013835
VERSION BE013835.1 GI:8274822
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keeler, U.W.
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred v0.960904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCA
BACKWARD: GTTTCACGTCACGACG
Plate: 52 row: B column: 23
Seq primer: ATTAGGTGACCTATAG.

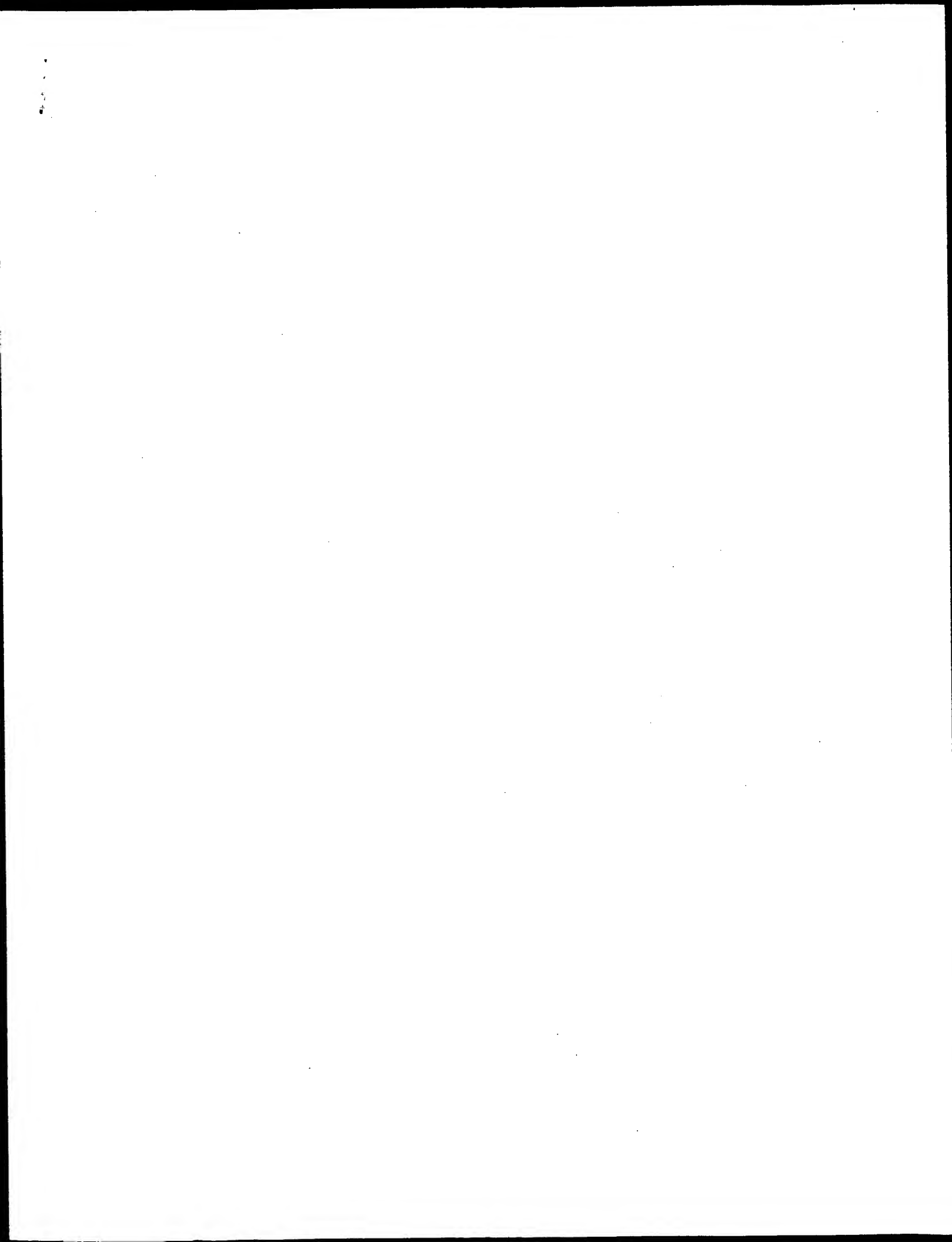
FEATURES

Location/Qualifiers
1..545
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPOR6; Site 1: XbaI; Site 2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."
BASE COUNT 128 a 120 c 129 g 168 t
ORIGIN

Query Match 0.9%; Score 74; DB 10; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4935 gttttgtagacaccttcagagatggtagagcttatcctactaagaagacttca 4994
|||||
DB 216 GTTTTGTAGACACCTTTTCAGATGGGTAGAGCTTATCCTACTAAGAAGACTTCA 157
|||||
QY 4995 accgtgtggtctaa 5008
|||||
DB 156 ACCGTGTGCTTAA 143
|||||

Search completed: February 24, 2002, 04:28:43
Job time: 24069 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 04:28:43 ; Search time 6170.56 Seconds
(without alignments)
5147.753 Million cell updates/sec

Title: US-09-171-553b-9

Perfect score: 2956

Sequence: 1 tgcctttaggggttagaac.....aaaaaaaaaaaaaaaaa 2956

Scoring table:

OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size: 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estol:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_iny:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	190	6.4	564	11	B1182930 UNL-P-FN-
C 2	92	3.1	558	11	B1360879 388957 MA
C 3	92	3.1	602	11	BF712162 MI-P-E6-a
C 4	78	2.6	549	10	AM657531 110922 MA
C 5	69	2.3	468	11	B1186066 UNL-P-FN-
C 6	69	2.3	468	11	BF704354 MI-P-O3-a
C 7	69	2.3	521	11	B1399852 MI-P-AV1-
C 8	69	2.3	730	11	B1185535 UNL-P-FN-
C 9	68	2.3	468	11	B1182727 UNL-P-FN-
C 10	64	2.2	485	11	BF703671 MI-P-E4-a
C 11	64	2.2	554	11	BF703805 MI-P-O2-a
C 12	52	1.8	451	13	A2891915 RPCI-24-1

C 13	51	1.7	545	11	BF701947 MI-P-E3-a
C 14	49	1.7	300	11	B1184279 UNL-P-FN-
C 15	48	1.6	494	11	B1182742 UNL-P-FN-
C 16	48	1.6	595	11	B1185465 UNL-P-FN-
C 17	47	1.6	294	11	BF713588 MI-P-O2-a
C 18	47	1.6	576	11	A2730616 RPCI-24-1
C 19	47	1.6	579	11	B1183723 UNL-P-FN-
C 20	40	1.4	166	10	AM446655 29171 MAR
C 21	40	1.4	631	11	BF702741 MI-P-E3-a
C 22	38	1.3	403	10	AM447645 89654 MAR
C 23	37	1.3	482	13	AQ920035 RPCI-23-2
C 24	36	1.2	545	10	AA529580 V41404.r
C 25	36	1.2	610	13	A2859488 2M0152P15
C 26	36	1.2	640	13	A2997740 2M0284R23
C 27	36	1.2	646	13	A2638811 1M0498123
C 28	36	1.2	655	13	A2740414 RPCI-24-1
C 29	36	1.2	660	13	A2353293 1M0092E07
C 30	35	1.2	673	13	A2329067 1M0053610
C 31	35	1.2	857	11	B1184013 UNL-P-FN-
C 32	34	1.2	155	10	AM446676 29231 MAR
C 33	34	1.2	261	11	B1184980 UNL-P-FN-
C 34	34	1.2	345	13	A2780588 2M0018A11
C 35	34	1.2	513	10	AM786160 119054 MA
C 36	34	1.2	550	13	A2390750 1M0152P16
C 37	34	1.2	574	13	A2390749 1M0152P15
C 38	32	1.1	340	13	BH068193 2M0236A21
C 39	32	1.1	348	13	A2966390 RPCI-24-2
C 40	31	1.0	251	11	BG382013 297673 MA
C 41	31	1.0	257	10	AA245675 mx05911.r
C 42	31	1.0	432	11	B1081943 602879172
C 43	31	1.0	834	13	BH124666 RPCI-24-2
C 44	30	1.0	114	11	BF728989 1000070E0
C 45	30	1.0	230	10	AM786010 117582 MA

ALIGNMENTS

RESULT 1
LOCUS B1182930 564 bp mRNA
DEFINITION UNL-P-FN-dr-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
B1182930
B1182930 UNL-P-FN-dr-h-08-0-UNL 3', mRNA sequence.

ACCESSION B1182930.1 GI:14657339

VERSION B1182930.1

KEYWORDS EST.

SOURCE pig.

ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 564)

AUTHORS Caetano, A.R., Johnson, R.K. and Pomp, D.

TITLE Generation and sequence characterization of a normalized cDNA

JOURNAL library from swine ovarian follicles

COMMENT Unpublished (2001)

CONTACT: Pomp, D

DEPARTMENT: Department of Animal Science

UNIVERSITY: University of Nebraska, Lincoln

ADDRESS: Lincoln, NE 68583-0908, USA

TELEPHONE: Tel: 402 472 6416

FAX: Fax: 402 472 6362

EMAIL: Email: dpomp@unl.edu

NOTES: The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized porcine ovarian follicles library

Seg primer: M13 -29

POLYA-yes.

Location/Qualifiers
1..564
/organism="Sus scrofa"

```

/strain="University of Nebraska, Lincoln Swine Selection
Lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-br-08-0-UNL"
/clone_1id="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pUT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F2. The tag is a string of 5-6
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996.
TAG_LIB=UNL-P-FN
TAG_TISSUE=porcine ovarian follicles
TAG_SEQ=CACACT"
141 a 129 c 143 g 151 t
BASE COUNT
ORIGIN

```

Query Match	6.4%: Score 190; DB 11; Length 564;
Best Local Similarity	99.3%: Pred. No. 5.4e-37;
Matches 290; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY 2467 cctgtgcgtccagtaataagtagtaagaagtcacacttcctattgttccaggccctgcta	2526
Db 564 CCTGTGCGCTCCAGTAATAATAGTAGAAGGTCACACTTCCTATGTGTCAGGGCCTGCTA	505
QY 2537 tctgtgcctaaagtaagataacagagaatagtgtaactaatgcgttatcttgatcttcgta	2586
Db 504 TCTGCGCCTTAAGTAGAAGTAACAGAAAGCACTTGCTATGCTTATGTGGAATTCGTAA	445
QY 2587 aactgcagtcgagcactatagaagaatattgattacacattgacagccctcagactatctaa	2646
Db 444 AACTGCAGTCGCAACCTAATAAGACATCGATTACACATGTGACAGCCCTAAGTACCTACTCA	385
QY 2647 ctgcacatctgtactactctgcagagagccacagcagatgcggaacctcggagctatttaa	2706
Db 384 CTGCATATCTGTACACTCTGCCAGAGGCCACAGCAATGCGGACCTCCAGACTATTTTAA	325
QY 2707 aatgatctgtccagagagcgcggtctcgtattttaaattgatgtgcca	2758
Db 324 AATGATTTGCTCCAGCGAGCGGGCTCTCGATATTTTTAAATGATTGTCTCA	273
RESULT 2	
LOCUS B1360879 558 bp mRNA EST 01-AUG-2001	
DEFINITION 386957 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.	
ACCESSION B1360879	
VERSION B1360879.1 GI:15056907	
KEYWORDS EST.	
SOURCE pig.	
ORGANISM Sus scrofa	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.	
REFERENCE 1 (bases 1 to 558)	
AUTHORS Fahrrekrug,S.C., Fieking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,	
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.	
and Keeler,J.W.	
Design and use of two pooled tissue normalized cDNA libraries for	
EST discovery in swine	
Unpublished (2000)	
CONTACT Contact: US Meat Animal Research Center	
USDA, ARS, Smith TPL	
PO Box 166, Clay Center, NE 68933-0166, USA	
Tel: 402 762 4366	
Fax: 402 762 4390	
TITLE	
JOURNAL	
COMMENT	

```

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCATATGACCACAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 142 row: B column: 4
Seq primer: ATTATGGACATCATATG.

FEATURES
    Location/Qualifiers
        1..558 - "BAC clone"
SOURCE

```

```

/ote="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI
Library made from pooled tissue from testis, ovary,
endometrium, hypochlamus, pituitary, and placenta."
BASE COUNT      129 a      123 c      119 g      187 t
ORIGIN

```

Query Match	3.1%;	Score 92;	DB 11;	length 558;
Best Local Similarity	100.0%;	Pred. No.	2.3e-13;	
Matches	92;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

[illegible]

RESULT	3
Bf712162/c	
LOCUS	Bf712162
DEFINITION	Bf712162 602 bp. mRNA EST 02-JAN-2001
ACCESSION	M1-P-E6-acx-e-11-1-UM.s1 MI-P-E6 Sus scrofa CDNA clone
VERSION	Bf712162
KEYWORDS	Bf712162.1 GI:12011595
SOURCE	EST.
ORGANISM	pig.
	Sus scrofa
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
TITLE	1 (bases 1 to 602) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9), 791-806 (1996)
JOURNAL	
MEDLINE	97044477
COMMENT	Contact: Tugale CK

201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuglie@astate.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized fetus at gestational day 20 library cDNA library preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
PolTA=yes.

FEATURES

source

Location/Qualifiers

1..602
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-E6-ack-e-11-1-UM"
/lab_host="MI-P-E6"
/lab="host=DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: EcoRI. The MI-P-E6 library is derived from fetus at gestational day 20. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.
TAG_LIB=MI-P-E6
TAG_TISSUE=fetus at gestational day 20
TAG_SEQ=AGGAA"

BASE COUNT 141 a 146 c 152 g 162 t 1 others
ORIGIN

Query Match 3.1%; Score 92; DB 11; Length 602;
Best Local Similarity 99.3%; Pred. No. 2.1e-13;
Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2807 ggccttggtggaacccataaagctgtccgattccgagctgggcccagctctct 2866
|||||
DB 143 ggcttgggtggaacccataaagctgtccgattccgagctgggcccagctctct 84
|||||

QY 2867 accctctgctgtgtgtaagctgtggccagcagcgttggaataaactctctgtc 2926
|||||
DB 83 accctctgctgtgtgtaagctgtggccagcagcgttggaataaactctctgtc 24
|||||

QY 2927 ttggcatcaaaaaaataaataa 2949
|||||
DB 23 ttggcatcaaaaaaataaataa 1
|||||

RESULT 4
AM657531 549 bp mRNA EST 25-APR-2001
LOCUS 110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AM657531
VERSION AM657531.1 GI:7423429
KEYWORDS EST.
SOURCE EST.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 549)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahnenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.W., Kohrer,G.A., Chitko-McKown,C.G.,
Perteira,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
Keeler,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 94 row: E column: 12

FEATURES

source

Seq primer: ATTTAGTGACCTATAG.
Location/Qualifiers

1..549
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

BASE COUNT 145 a 144 c 121 g 139 t
ORIGIN

Query Match 2.6%; Score 78; DB 10; Length 549;
Best Local Similarity 99.2%; Pred. No. 5.4e-10;
Matches 128; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1387 aaatcaatgtgacccaagacatagaatcaatgtacatgggtcccgaaataagctt 1446
|||||
DB 159 aaatcaatgtgacccaagacatagaatcaatgtacatgggtcccgaaataagctt 218
|||||

QY 1447 acccttaactgaagtttcgggaaggagacatgacatagaagaagctcccatccaccaa 1506
|||||
DB 219 acccttaactgaagtttcgggaaggagacatgacatagaagaagctcccatccaccaa 278
|||||

QY 1507 caacttttc 1515
|||||
DB 279 caacttttc 287
|||||

RESULT 5
B1186066 468 bp mRNA EST 10-JUL-2001
LOCUS B1186066/c
DEFINITION UNL-P-FN-bo-a-05-0-UNL.s2 UNL-P-FN Sus scrofa cDNA clone
ACCESSION B1186066
VERSION B1186066.1 GI:14660475
KEYWORDS EST.
SOURCE EST.
ORGANISM pig.
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 468)
Caetano,A.R., Johnson,R.K. and Pomp,D.
Generation and sequence characterization of a normalized cDNA
library from swine ovarian follicles
unpublished (2001)
Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu

TITLE The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail.
JOURNAL
COMMENT

Seq primer: M13 -29
POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..468
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection
lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-bo-a-05-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.

BASE COUNT 127 a 98 c 97 g 146 t
TAG_SEQ=None found"

ORIGIN

Query Match 2.3%; Score 69; DB 11; Length 468;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2884 gactgtggcccgccgcttggaataaatacctctgtctgttcgcatcaaaaaa 2943
|||||
Db 69 GACTGTGGCCCGCCGCTTGGAATAAATCCTCTGCTTTCGATCAAAAAA 10
QY 2944 aaaaaaaa 2952
|||||
Db 9 AAAAAAAA 1

RESULT 6
BF704354/c 499 bp mRNA EST 22-DEC-2000
LOCUS
DEFINITION MI-P-03-aba-g-01-1-UM.s1 MI-P-03 Sus scrofa cDNA clone
ACCESSION BF704354
VERSION MI-P-03-aba-g-01-1-UM 3', mRNA sequence.
KEYWORDS BF704354.1 GI:11989756
EST.
SOURCE
ORGANISM pig.
Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 499)
TITLE Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tugale CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktugale@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: RJ Woods, JA Green, KS Praher, S142
Animal Science Research Center, Department of Animal Science,
University of Missouri-Columbia, 65211 Clone distribution: clones
will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
Source Location/Qualifiers

1. 499
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-03-aba-g-01-1-UM"
/clone_lib="MI-P-03"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-03
library is derived from ovary at estrus day 12. For a

detailed description of the library from which this clone
was derived, please visit our web site at
<http://pigst.genome.iastate.edu/>.
TAG_SEQ=None found"

BASE COUNT 132 a 104 c 107 g 152 t 4 others

Query Match 2.3%; Score 69; DB 11; Length 499;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;
Matches 69; Conservative 0; Mismatches 0; Indels 0;

QY 2884 gactgtggcccgccgcttggaataaatacctctgtctgttcgcatcaaaaaa 2943
|||||
Db 69 GACTGTGGCCCGCCGCTTGGAATAAATCCTCTGCTTTCGATCAAAAAA 10
QY 2944 aaaaaaaa 2952
|||||
Db 9 AAAAAAAA 1

RESULT 7
B1399852/c 521 bp mRNA EST 14-AUG-2001
LOCUS
DEFINITION MI-P-AV1-nrv-d-03-0-UI.s1 MI-P-AV1 Sus scrofa cDNA clone
ACCESSION B1399852
VERSION MI-P-AV1-nrv-d-03-0-UI 3', mRNA sequence.
KEYWORDS B1399852.1 GI:15178913
EST.
SOURCE
ORGANISM pig.
Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 521)
TITLE Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tugale CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktugale@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized placenta library cDNA library Preparation: M.B. Soares
lab, University of Iowa EST sequencing: M.B. Soares lab, University
of Iowa Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
Source Location/Qualifiers

1. 521
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-AV1-nrv-d-03-0-UI"
/clone_lib="MI-P-AV1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AV1
library is normalized library derived from the MI-P-AV1
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
<http://pigst.genome.iastate.edu/>. The procedure used to
create this library has been previously described (Bonaldo

, Lennon and Soares, Genome Research 6: 791-806, 1996)
 TAG_LIB=M1-P-AVL
 TAG_TISSUE=placenta
 TAG_SEQ=ATCG"

BASE COUNT 142 a 106 c 103 g 170 t
 ORIGIN
 Query Match 2.3%; Score 69; DB 11; Length 521;
 Best Local Similarity 100.0%; Pred. No. 8.4e-08;
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2884 gactgtggcccgccgcttggaataaaatcctctgctgttgcacataaaaaa 2943
 Db 69 GACTGTGGGCCCGCCGCTTGGAATAAAATCCTCTGCTTGCATCAAAAAA 10

QY 2944 aaaaaaaaaa 2952
 Db 9 AAAAAAAAAA 1

RESULT 8
 BI185535
 LOCUS BI185535 730 bp mRNA EST 10-JUL-2001
 DEFINITION UNL-P-FN-cx-a-05-0-UNL.s1 UNL-P-FN Sus scrofa CDNA clone
 ACCESSION UNL-P-FN-cx-a-05-0-UNL 3', mRNA sequence.
 VERSION BI185535
 KEYWORDS BI185535.1 GI:14659944
 SOURCE EST.
 ORGANISM Pig.
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 730)
 AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 TITLE Caetano, A.R., Johnson, R.K. and Pomp, D.
 JOURNAL Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 COMMENT Unpublished (2001)
 Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu
 Oligo-dt track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message.
 Seq primer: M13 -29
 POLYA=No.

FEATURES

source

Location/Qualifiers
 1..730
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection
 Lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-cx-a-05-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pTR3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
 library is a normalized library representing porcine
 ovarian follicles, ranging between 2.0 to 10.0 mm in
 diameter, collected during 7 days of the follicular phase
 of the pig estrous cycle. This library was derived from
 the library UNL-P-F2. The tag is a string of 5-6
 nucleotides present between the Not I site and the
 oligo-dt track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996.
 TAG_SEQ=None found"

BASE COUNT 244 a 149 c 161 g 176 t
 ORIGIN

Query Match 2.3%; Score 69; DB 11; Length 730;
 Best Local Similarity 100.0%; Pred. No. 6.1e-08;
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2374 aagaagtgggaatgaaatgaatgaatgaatgaatgaatgaatgaatga 2433
 Db 649 AAGAAGTGGGAATGAAAGATGAATGAATGAATGAATGAATGAATGAATGA 708

QY 2434 aaaaaaaaaa 2442
 Db 709 AAAAAAAAAA 717

RESULT 9
 BI182727/c
 LOCUS BI182727 468 bp mRNA EST 10-JUL-2001
 DEFINITION UNL-P-FN-bo-a-05-0-UNL.s1 UNL-P-FN Sus scrofa CDNA clone
 ACCESSION UNL-P-FN-bo-a-05-0-UNL 3', mRNA sequence.
 VERSION BI182727
 KEYWORDS BI182727.1 GI:14657136
 SOURCE EST.
 ORGANISM Pig.
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 468)
 AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 TITLE Caetano, A.R., Johnson, R.K. and Pomp, D.
 JOURNAL Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 COMMENT Unpublished (2001)
 Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu

The sequence contained an oligo-dt track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail.
 Seq primer: M13 -29
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..468
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection
 Lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-bo-a-05-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pTR3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
 library is a normalized library representing porcine
 ovarian follicles, ranging between 2.0 to 10.0 mm in
 diameter, collected during 7 days of the follicular phase
 of the pig estrous cycle. This library was derived from
 the library UNL-P-F2. The tag is a string of 5-6
 nucleotides present between the Not I site and the
 oligo-dt track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996.
 TAG_SEQ=None found"

BASE COUNT 127 a 98 c 97 g 145 t 1 others
 ORIGIN
 Query Match 2.3%; Score 68; DB 11; Length 468;
 Best Local Similarity 100.0%; Pred. No. 1.6e-07;

Matches 68: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2884 gacgttgagcccgagcgttggaataaataacgtctgtctgttcacataaaaaa 2943
 |||||||
 Db 68 GACGTGTGGCCCCCAGCCGCTTGGAATAAATCTCTCTGTTCATCAAAAAA 9
 |||||||

OY 2944 aaaaaaa 2951
 |||||||
 Db 8 AAAAAA 1

RESULT 10
 BF703671/c 485 bp mRNA EST 22-DEC-2000
 LOCUS MI-P-E4-abp-c-04-1-UM.s1 MI-P-E4 Sus scrofa cDNA clone
 DEFINITION MI-P-E4-abp-c-04-1-UM 3', mRNA sequence.
 ACCESSION BF703671
 VERSION BF703671.1 GI:11989079
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 485)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ctuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized embryo at gestational day 14 library cDNA library preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

Location/Qualifiers

1..485
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-E4-abp-c-04-1-UM"
 /clone_lib="MI-P-E4"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E4 library is derived from embryo at gestational day 14. For a detailed description of the library from which this clone was derived, please visit our web site at http://piglet.genome.iastate.edu/
 TAG_LIB=MI-P-E4
 TAG_TISSUE=embryo at gestational day 14
 TAG_SEQ=ACTCAC"
 BASE COUNT 142 a 114 c 109 g 119 t 1 others
 ORIGIN

Query Match 2.2%; Score 64; DB 11; Length 485;
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2819 aaccataaaagctgtccgacttcgcagctcgggcccagctctaccctcgtg 2878
 |||||||
 Db 131 AACCCATTAAGCTGTCCGATTCGCCAGCTCGGGGCGCACTCTCAACCCCTCG 72
 |||||||

OY 2879 tgta 2882
 |||||
 Db 71 TGTA 68

RESULT 11
 BF703805/c 554 bp mRNA EST 22-DEC-2000
 LOCUS MI-P-O2-abs-a-11-1-UM.s1 MI-P-O2 Sus scrofa cDNA clone
 DEFINITION MI-P-O2-abs-a-11-1-UM 3', mRNA sequence.
 ACCESSION BF703805
 VERSION BF703805.1 GI:11989213
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 554)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ctuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized ovary at estrus day 5 library cDNA library preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

Location/Qualifiers

1..554
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-O2-abs-a-11-1-UM"
 /clone_lib="MI-P-O2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-O2 library is derived from ovary at estrus day 5. For a detailed description of the library from which this clone was derived, please visit our web site at http://piglet.genome.iastate.edu/
 TAG_LIB=MI-P-O2
 TAG_TISSUE=ovary at estrus day 5
 TAG_SEQ=ATCATC"
 BASE COUNT 158 a 128 c 122 g 146 t
 ORIGIN

Query Match 2.2%; Score 64; DB 11; Length 554;
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2819 aaccataaaagctgtccgacttcgcagctcgggcccagctctaccctcgtg 2878
 |||||||

DB	134	AA0CCCAATAAAAG6GTGTCCCGAATTCGGCACTCGGGGGCCGAGTCCCTCTCTACCCCTCGGTGG	75
OY	2879	tgtA 2882	
Db	74	TCTA 71	
RESULT	12		
AZ891915/C			
LOCUS			
DEFINITION	AZ891915	451 bp	DNA
	RPCT-24-172A20.TV	RPCT-24	Mus musculus genomic clone RPCT-24-172A20
ACCESSION			
VERSION	AZ891915		
KEYWORDS	AZ891915.1	GT:13210860	
SOURCE			
ORGANISM	house mouse.		
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 451)		
	Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akınret, B., Levins, M.,		
	Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorgis, E.,		
	Russell, D., de Jong, P. and Fraser, C. M.		
	Mouse BAC End Sequences from Library RPCT-24		
	Unpublished (1999)		
	Other_GSSs: RPCT-24-172A20.TU		

```

Department of Eukaryotic Genomics
The Institute for Genomic Sciences
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCT-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@email.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 172 row: A column: 20
Seq primer: 17
Class: BAC ends.
Location/Qualifiers
1. .451
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCT-24-172A20"
/clone_lib="RPCT-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTRBAC1; site_1: BamHI; site_2: BamHI;
RPCT-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTRBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT
144 a 91 c 106 g 110 t
ORIGIN
```

[illegible]

ACCESSION	BF701947
VERSION	BF701947.1
KEYWORDS	GI:11987355
SOURCE	EST.
ORGANISM	pig.
	Sus scrofa
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
AUTHORS	1 (bases 1 to 545)
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B.
	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Tuggle CK Molecular Genetics Laboratory, Department of Animal Science Iowa State University 201 Kildee Hall, Ames, IA 50011-3150, USA Tel: 5152944252 Fax: 5152942401 Email: cktuggle@iastate.edu The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized fetus at gestational day 45 library cDNA library Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com) Seq primer: M13 forward POLY-A=yes.
FEATURES	Location/Qualifiers
Source	I. 345

		/strain="crossbred"
		/db_xref="taxon:9823"
		/clone="MI-P-E3-aag-c-06-1-UM"
		/clone_lib="MI-P-E3"
		/lab_host="DH10B (Life Technologies)"
		/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E3 library is derived from fetus at gestational day 45. For a detailed description of the library from which this clone was derived, please visit our web site at http://pigest.genome.iastate.edu/ .
		TAG_LIB=MI-P-E3
		TAG_TISSUE=fetus at gestational day 45
BASIS COUNT	138 a	120 c
ORIGIN		121 g 166 t
		TAG_SEQ=CACGAC"
Query Match	1.7% ; Score 51; DB 11; Length 545;	
Best Local Similarity	100.0%; Pred. No. 0.0018;	
Matches	51; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Db	51 GCCTGGAAATVAAAAATCCTTCGTGCTTGATCAAAAAAAAAAAAAA 1	
QY	2902 gcctgaataaaaaaacctccttgcgtgctttggcatcaaaaaaaaaaaa 2952 	
RESULT 14		
BI184279/c		
LOCUS	300 bp mRNA EST	10-JUL-2001
DEFINITION	UNT.-P-FN-ck-g-02-0-UNT.s1 UNT.-P-FN Sus scrofa CDNA clone	
ACCESSION	BI184279	
VERSION	BI184279.1 GI:14658688	
KEYWORDS	EST.	
SOURCE	pig	
ORGANISM	Sus scrofa	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

REFERENCE 1 (bases 1 to 300)
 AUTHORS Caetano, A.R., Johnson, R.K. and Pomp, D.
 TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu

The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail.

Seq primer: M13 -29
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..300
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-ck-g-02-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DHI0B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
 TAG_SEQ=None found"

BASE COUNT 69 a 62 c 71 g 97 t 1 others
 ORIGIN

Query Match 1.7%; Score 49; DB 11; Length 300;
 Best Local Similarity 100.0%; Pred. No. 0.0094;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2884 gactgtggcccgagcgttgaaataaacctctgtgttgcga 2932
 ||||||||||||||||||||||||||||||||||||||||
 DB 66 GACTGTGGCCCGAGCGCTTGGAATAAATCCTTGTGATTCGA 18

RESULT 15
 LOCUS B1182742 494 bp mRNA EST 10-JUL-2001
 DEFINITION UNL-P-FN-bo-e-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 ACCESSION B1182742
 VERSION B1182742.1 GI:14657151
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 494)
 AUTHORS Caetano, A.R., Johnson, R.K. and Pomp, D.
 TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science

University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu
 The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the normalized porcine ovarian follicles library
 Seq primer: M13 -29
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..494
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-bo-e-05-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DHI0B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
 TAG_LIB=UNL-P-FN
 TAG_TRISUP=porcine ovarian follicles
 TAG_SEQ=CACTC"

BASE COUNT 118 a 117 c 125 g 134 t
 ORIGIN

Query Match 1.6%; Score 48; DB 11; Length 494;
 Best Local Similarity 100.0%; Pred. No. 0.01;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2889 tggcccgagcgcgttgaaataaacctctgtgttgcacaa 2936
 ||||||||||||||||||||||||||||||||||||||||
 DB 64 TGGGCCCGAGCGCGCTTGGAATAAATCCTTGTGATTCGA 17

Search completed: February 24, 2002, 04:28:52
 Job time: 24078 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 20:59:59 ; Search time 6550.43 Seconds

(without alignments)
5446.358 Million cell updates/sec

Title: US-09-171-553B-1
Perfect score: 3320
Sequence: 1 gaattcgccgcgcgcgcac.....aaaaaaaaaaaaaaaaaa 3320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	385	11.6	576	13	A2730616
2	380.8	11.5	864	13	BH056750
3	373.2	11.2	673	13	A2329067
4	348.8	10.5	799	11	B1107577
5	348.8	10.5	705	11	A2337037
6	341.6	10.3	842	13	BH106844
7	340.2	10.2	789	13	A2715956
8	338.8	10.1	755	13	A2715272
9	336.8	10.1	653	13	BH083773
10	335	10.1	700	13	A2827524
11	333.2	10.0	689	13	A2897994
12	331.4	10.0	817	13	A2717731

C 13	330.8	10.0	692	13	A2759626	A2759626
C 14	328	9.9	734	13	BH074726	BH074726
C 15	327.6	9.9	610	13	A2432720	A2432720
C 16	326.6	9.8	685	13	A2613787	A2613787
C 17	324.6	9.8	651	13	A2940771	A2940771
C 18	324	9.8	642	13	A2624191	A2624191
C 19	323	9.7	616	13	A2962118	A2962118
C 20	322	9.7	656	13	A2560266	A2560266
C 21	320.6	9.7	652	13	A2758880	A2758880
C 22	319.2	9.6	725	13	A2866393	A2866393
C 23	318.4	9.6	668	13	A2960291	A2960291
C 24	316.8	9.5	681	13	A2466041	A2466041
C 25	316	9.5	618	13	BH044907	BH044907
C 26	314.6	9.5	646	13	BH041228	BH041228
C 27	313.8	9.5	639	13	A2643009	A2643009
C 28	313.2	9.4	701	13	A2860071	A2860071
C 29	313	9.4	702	13	A2974419	A2974419
C 30	312.6	9.4	680	13	A2747906	A2747906
C 31	310.8	9.4	575	13	A2479717	A2479717
C 32	310.6	9.4	665	13	A2377711	A2377711
C 33	310.4	9.3	656	13	A2848987	A2848987
C 34	309.6	9.3	675	13	A2915310	A2915310
C 35	309.4	9.3	688	13	A2825369	A2825369
C 36	308.6	9.3	600	13	A2987182	A2987182
C 37	308.4	9.3	691	13	BH034484	BH034484
C 38	308.2	9.3	568	13	A2824214	A2824214
C 39	308.2	9.3	638	13	A2435891	A2435891
C 40	307.8	9.3	630	13	A2596966	A2596966
C 41	306.4	9.2	640	13	A2867739	A2867739
C 42	306.2	9.2	626	11	BG065719	BG065719
C 43	304.8	9.2	630	13	A2838774	A2838774
C 44	303.4	9.1	586	13	A2416244	A2416244
C 45	302.8	9.1	583	13	A2415280	A2415280

ALIGNMENTS

RESULT 1					
LOCUS	A2730616	576 bp	DNA	GSS	25-JAN-2001
DEFINITION	RPCT-24-149D1.TJ RPCT-24 Mus musculus genomic clone RPCT-24-149D1.				
ACCESSION	A2730616				
VERSION	A2730616.1				
KEYWORDS	GI:12491859				
SOURCE	GSS.				
ORGANISM	house mouse.				
	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus 1 (Phases 1 to 576)				
AUTHORS	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinnit,B., Levins,M., Tsengaye,G., Geer,K., Kroll,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.				
TITLE	Mouse BAC End Sequences from Library RPCT-24				
JOURNAL	Unpublished (1999)				
COMMENT	Other_GSSs: RPCT-24-149D1.TJ				

FEATURES

Query Match	11.6%;	Score 385;	DB 13;	length 576;
Best Local Similarity	82.5%;	Pred. NO. 1.6e-66;		
Matches 453; Conservative	0;	Mismatches 95;	Indels 1;	Gaps 1;

REFERENCE	TITLE	JOURNAL	COMMENT
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.			
1 (bases 1 to 864)			
Zhao S., Nieman, W., Malek, J., Shatsman, S., Akinet, B., Iewins, M.,			
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,			
Russell, D., de Jong, P., and Fraser, C.M.			
Mouse BAC End Sequences from Library RPCI-24			
Unpublished (1999)			
Other GSSs: RPCI-24-94C16.TJ RPCI-24-94C16.TV			
Contact: Shaying Zhao			
Department of Eukaryotic Genomics			
The Institute for Genomic Research			
9712 Medical Center Dr., Rockville, MD 20850, USA			
Tel: 301 838 0200			
Fax: 301 838 0208			
Email: szhaot@igrr.org			
Clones are derived from the mouse BAC library RPCI-24. For BAC			
library availability, please contact Pieter de Jong			
(pdejong@mail.cno.org). Clones may be purchased from BACPAC			
Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end			
page: http://www.tlgr.org/tlgr/bac_ends/mouse/bac_end_intro.html			
Plate: 94 row: C column: 16			
Seq primer: T7			
Class: BAC ends.			
FEATURES			
Source	Location/Qualifiers		
	1..864		
	/organism="Mus musculus"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="RPCI-24-94C16"		
	/clone_id="RPCI-24"		
	/sex="Male"		
	/cell_type="Spleen/Brain"		
	/note="Vector: pTARPA1; Site_1: BamHI; Site_2: BamHI;		
	RPCI-24 Mouse BAC library produced by Pieter de Jong. The		
	library was cloned in the pTARPA1 cloning vector at the		
	BamHI sites using MboI partially digested male C57BL/6J		
	DNA."		
BASE COUNT	237 a 221 c 209 g 197 t		
ORIGIN			
Query Match	11.5%: Score 380.8; DB 13; Length 864;		
Best Local Similarity	66.5%: Pred. No. 1e-65;		
Matches 577; Conservative 0; Mismatches 287; Indels 4; Gaps 2;			
QY 463 gtgagagattttggygaagctgtgatttcgacagctgtgatacccgagtttgagacc 522			
Db 1 GTACGTGAGTTTGGGACAGCGAGCGTTTGTGACTGTGAATACAGAGGTCCACAAATG 60			
QY 523 ttgacagcccccacactctcccgctaaccaagaagaaagggagattctctctgagctctgag 562			
Db 61 TTACGGCGACCTCTGTGCCACACTACATAAAGAAAAAATCCCGTWCACCTGGACAGAGAA 120			
QY 583 caaccagaagcatttgatgcatcatcaaaaagccctgtctgagcgaccctgtcgccctc 642			
Db 121 CACCAAAAAGACTTTAACAGCATTAAGACGTCGCTCTGTCGACGCCAGCGCTTGGCCCTG 180			
QY 643 cctgagctaacctaaacccttacccttattatgtgatatgacgttaagggatagccgagga 702			
Db 181 CCAATTATTGACTAAGCCTTTACACCATTTAGCTGATGATGACGCACCTGGAGTAGCCCGCAGA 240			
QY 703 gtttaaccccaaacccaggacacataggagagagacocytgttcttaccactgtaagaagctt 762			
Db 241 GTCTGTACTAGACTCTGGGGCGGTGAAAGAGGTATGTGCTTAACCTGTCTAAAAAATTG 300			
QY 763 gatccctttagccagttgttggccgctatgtcgtgaagctatacgcagctgtgacatactg 822			
Db 301 GACCCAGTGTCCAGTGAATGCGCTCTGCTTAAAGCCATTCGCTCAGTAGGCCCTACTT 360			
QY 823 gtcaaggaagcgtacaaattgtgactttgggacagaaataaactgtaataagcccccatgca 882			
Db 361 GTTAAAGATATCTGGCAAGCTTACTCTAAGGTCACACATGTGATGTATCACTCCCATATGGC 420			

QY	883	tlgagagaactcgtlttgggaagcccccaagccgatlgatataccaagcccgcatgacac	942
Db	421	TTAAAGAGTATTGATATACAGCCCCCTGCACCATTTGATGACAAATATCCCGATGACAC	480
QY	943	latcaagagcgtccttcacacagagaggtgaacttlgtccacacagccgtctcaacct	1002
Db	481	TACACGAGCCTGTTCCTAATATGACCGAGTGACTTTGGCCCTACCTCCATCCACAACTCT	540
QY	1003	gcacctcttctgccttgaaagagactgatataacagtgacttcattgatitgcataactat	1062
Db	541	ACCCACCCCTTCCCCACAGACGATGACTTACCTGGGTACACCAATGTCTGATATCCTG	600
QY	1063	atlaagagagcttggtgtccgcaagagaccttaacagataccgcttgactggaagatgcta	1122
Db	601	GATTAAGAAACTGGATTCGAGAAAGATCTGACTCAGCAACAACTTGGGC---AGGTGGCT	657
QY	1123	acctgtctcactgaacggaagcagctatgtgtgtgaaagttaagaagatgtcgttggcgcta	1182
Db	658	GACAGGTACACAGATGGTATGACAGCTTGTGCTGGAGGGGAAATAAGACACAGAGGGA	717
QY	1183	gttgatgaacggagaccgcagactgctgtggccagcagccttgcgcgaagaagacttaagcgaa	1242
Db	718	GTGGTGCACAAAACATGTAATTTTGGGCAAGTTAACTTCCAAAGAACCTTCCGCTCAG	777
QY	1243	aagcgtgagctatgatgcgccttaacgcaagccttgcggtgtgccgaaggaatccataaac	1302
Db	778	AAGGCTGAGCTCTGACGCTAACATAAAGACTCTGCCTGTG-CAAAAGGTATGAGCAATCAAT	836
QY	1303	altatatacggacacagatagctttg	1330
Db	837	ATTTAACTGAAGCACACTATGCTTTTG	864

```

RESULT      3
LOCUS       A2329067/c
DEFINITION  A2329067 673 bp DNA
            1M0053G10F Mouse 10kb plasmid UGCG1M Library
            A2329067 clone UGCG1M0053G10 F, DNA sequence.
ACCESSION   A2329067
VERSION     A2329067.1 GI:10389414
KEYWORDS    GSS.
SOURCE      house mouse.
            Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
REFERENCE   1 (bases 1 to 673)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0053 row: G column: 10
            Seq primer: CGTTGTAAACGACGCGCAGT
            Class: plasmid ends
            High quality sequence stop: 673.
FEATURES
            Location/Qualifiers
                1..673

```

```
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db_xref="taxon:10090"  
/clone="U06C1M0053G10"  
/clone_lib="Mouse 10kb plasmid U06C1M library"
```

/sex="Male"
 /lab host="E. Coli strain XL10-Gold, T1-resistant, F"
 /note=Vector: PWD42hv: Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 digested DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (g11473211419b|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

Query Match	Similarity	11.2%	Score 373.2	DB 13	Length 673
Best Local	Similarity	78.1%	Pred. No. 3,5e-64		
Matches	Conservative	0	Mismatches 123	Indels	Gaps
QY 1900	ctaaagggaagaccaccgagcgctcactctggaagtggacttcaactgaaglaaacgcgcgt	1959			
DB 672	cttcagcgatggctgccttgaaagccttacttgagggtagatgttaccgaatcctgaacgct	613			
QY 1960	aaatacggaaaataatcattctgttttttgaacaccttcaagatgggtlaagagct	2019			
DB 612	aaattacggaaacaaagtaacttattttagtattatagataccttttcaagggtgaatgaagct	553			
QY 2020	tatctactaagaagaagacttcaacgcgtgtgtgttaagaanaaatacttgaagaaatttt	2079			
DB 552	ttttcttaccacgacgtgacgac-----ggccaattatgattccttaacgcaaaattcgc	505			
QY 2080	ccaagatttgaatactactgaagttaataaggtcaagaatgggtccagcttctgtgccag	2139			
DB 504	ccgagctttgtatattaccacaggttaaacaggtcccaataatggacttccctttgttggccacg	445			
QY 2140	gtaagtcaggaagcttgcaccaagatattgggatttgcatttgaanaactgcattgttgatacaga	2199			
DB 444	gttaagtcaggaagcttgcaccagatatttagaatttgatttagaagttacatttgcattgcatacag	385			
QY 2200	ccccaaagctcaagagacaggtagaagagaatgaatagaacattaaagaaccccttaccaaa	2259			
DB 384	ccccaaagctcagagacaggtagagaaacgaatgaagaaacattt--aaagagaccttaccaaa	326			
QY 2260	tttagacagagaagacttgcatttaattgaattgtagtgcctccctgcaccttgcatttlaag	2319			
DB 325	tttggcagcttagagactgaagcttaattgattgaattgagcttccctaccccttggctcttcag	266			
QY 2320	gttagagaacaaccccttgaacagtttggcgcttgaacccctatgaattgtcctcac-----ggggg	2375			
DB 265	gtttagaacaacattcccttaaacagtttggacgtacacccctatataattatcttcaacggggggg	206			
QY 2376	accccccccgcttggcagaatattgccttgcattcacalaagtgcgatgtgtgtcttctccacgc	2435			
DB 205	gggggcttccactgttaaaaattacgctctgtacatattgctgcacatctgctgtttcccgacc	146			
QY 2436	tttgtttcttaagcttcaagcgctcgagtgtgtgtgaagcgaga-----gggtgaagcaagc	2491			
DB 145	ttttgttcttaagcttcaagcgctcgagtgtgtgtgaagcaagtgatgggcattggaagagac	86			
QY 2492	ttcggagagagctctctcagag--gagacttgcgaattccacatcgcttccaagtgttga	2549			
DB 85	ttccagaagggcttactcagaagaaagacgctgcgaaccccccacatcttcttacaagttataga	26			

QY 2550 ttcagt 2555
 Db 25 TTCGGT 20

RESULT 4
 LOCUS B1107577 799 bp mRNA EST 26-JUN-2001
 DEFINITION 602894023F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5039044 5',
 mRNA sequence.
 ACCESSION B1107577
 VERSION B1107577.1 GI:14558470
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 799)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9abs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M1107 row: n column: 05
 High quality sequence stop: 743.
 Location/Qualifiers
 1. 799

FEATURES
 source
 /organism="Mus musculus"
 /strain="C57BL/6J (f1)"
 /db_xref="taxon:10090"
 /clone="IMAGE:5039044"
 /clone_1lb="NCI_CGAP_Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 /stem_cell_origin=""
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-Sport6; Site: 1: SalI;
 Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 211 a 215 c 218 g 154 t 1 others

ORIGIN

Query Match 10.5%; Score 349; DB 11; Length 799;
 Best Local Similarity 70.6%; Pred. No. 2.2e-59;
 Matches 537; Conservative 0; Mismatches 215; Indels 9; Gaps 5;

QY 729 gaagagactgttctactgtcaagaagcttgatctgtgacgaagtggtgacctg 788
 Db 19 GCGGGCCCGGCTAGCTTACTATCAAGAACTGGACGCTGTGGCCAGTGGAGCCCTC 78

QY 789 atgtctgaaggtatcgacgtgtgtgacactgtgtcaaggaagcgtgacaattgactt 848
 Db 79 CTGCTCGCGAGCATGACGACGACGCTGTGAGTACAGATGCGGCAAACTGACTAT 138

QY 849 ggaagacataactgtatagcccccatgcatgtgagaacatcgttcggagcccc 908
 Db 139 GGGCCAGAAATGTTACTATATGTGGCCCAACATCTCTTGAGAGCATCATCAGCAACACC 198

QY 909 aagaccgatgatgaccaaagcccgatgacccactataaagccgcttcctcaagaag 968
 Db 199 GGACCGCTGGATGACCAAGCCCGAATGACGACATCAAGACCTATTGCTGACAGACG 258

QY 969 ggttaacttcgtccacacagccgctctcaaaccttccactcttctcgtgaagaactga 1028
 Db 259 AGTAGATTTCACCCCGCAGCATTCACACCCCGCTCTTACTACTCT---GAGGCTGA 315

QY 1029 tgaaccagttactcattgattgcatcaactattgattgaggagacttgggtccgcaagga 1088
 Db 316 CGAGGCCCTGTCACATAGTGTGAGAAATATGCGAGAGAGATGAGATCCGGCCAGA 375

QY 1089 ccttacagacataccgttgacttgagaagatgctaaccttgatctacgaggaagacgcta 1148
 Db 376 CTTACAGACCAACCTTGGCCAGGCGGATG---ACTGCTTCACGACGGAAGCGCTT 432

QY 1149 tctgtgtgaaggttaagaagatgagctgggtgagcagtggttgagcagaccgc-acagatc 1207
 Db 433 TGTGTGAGAGGTAAAGCGAAGGCTGGGCTGCAGTAGTAGAGTAAGCAAGGCTGATAT 492

QY 1208 gggccagacagctgcccgggaaggaacttcagcgaagaagctgagctcagccctcaagc 1267
 Db 493 GGGCCAGCACTGTGCGGAGGATGATCACTCAAAACCGGAATCAATCGCATTAATTC 552

QY 1268 aagcttgcgctgcgcggaaggaatccataacattatacagacacaggtatgctt 1327
 Db 553 AAGCTTTAAGGCTGGCGAGAAAGAGGCTCTTAATGTCTATACGACGCTGTACGCTT 612

QY 1328 ttgcgactgcacagctacacggggccatctataaacaaggggtgttactcagcag 1387
 Db 613 TTGCCAGGCTCATGTTTCACGAGCAATATACCGACACCGTGACGCTGACGCTGACG 672

QY 1388 ggaaggaaataaagaaca-aagaagaaattcctaagctctatagaagcct-tacattgcc 1445
 Db 673 GCAAAAGATATCAAAATATACAGAAAGAAATTCACCTTAATTAAGAGCCGTTCATTGCGC 732

QY 1446 aaaaagctagctatatacactgtctcctgagacataagaag 1486
 Db 733 CGGTAGGTGGCAATTATTCATTGCCAGGACCAACCAAGAG 773

RESULT 5
 A2337037 705 bp DNA GSS 29-SEP-2000
 LOCUS 1M0067L03R Mouse 10kb plasmid U08C1M library Mus musculus genomic
 DEFINITION clone U08C1M0067L03 R, DNA sequence.
 ACCESSION A2337037
 VERSION A2337037.1 GI:10406938
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 705)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 CONTACT: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel.: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0067 row: 1 column: 03
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 705.
 Location/Qualifiers
 1. 705

FEATURES
 source
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U08C1M0067L03"

Qy	1509	gaccacatgctgaccgggttccaaacagagacccagagctgttaaccttcgcat	1568
	721	CCACCGCATGCAACATCTACTAGCAAAACAAGTGGCTCAGAGGCCATATCTTACTGA	780
Db	1569	ataag	1573
	781	AAAAG	785

[illegible]

BASE COUNT	202 a	194 c	190 g	169 t
ORIGIN				

Query Match	10.2%;	Score 338.8;	DB 13;	Length 755;
Best Local Similarity	68.1%;	Pred. No. 2.3e-57;		
Matches 501; Conservative	0;	Mismatches 232;	Indels 3;	Gaps 2

Oy	49	ccccactgacacacccatttgccttcgaatgagagatccaggtacgagggaacacccgg	108
Db	23	cccaacaccaccccttcttgccttggaaatgagaaactccagaaatggacaaacccga	82
Oy	109	cagctacactgacccgaactcctcccaaggtctcaagaactcccgaccactcttgacga	168
Db	83	cagctacactgacacagacctctccacaggaattcaagaactcggccactttgttcgaa	142

OY	169	gctcaacaaaggagccctggccaactcgaagtaccacaacctcaagtgtaacctccag	228
Db	143	GCCCTACACGAGATCTTGCTCTTTTCCAAGCACAATAACCACAGTGATTCCTGTGA	202
OY	229	tacgtgatgaacctgctctctgcgcggagcccacaaacagaactcgttagaagtcaag	288
Db	203	TATGTAGATGACTCGTCTCTTAGCTGCAGAAACACAGTGGAGCTGTGAATTGGGACC	262
OY	289	gcactactgcctggaattgtctcgactaggtcacagagcctctgttaagaagcccaat	348
Db	263	AACCTCCTGGGCGGTTAAGTAAGCTGGGGATGCGGCTCTTGCTAAAAGAAGCTAG	322
OY	349	tgcagaaagagataaacactcttgggtacagtttcgcgggcggcagacgaatgctacg	408
Db	323	TGCCAGATAAAAAGTGACTACTACCTAGAGATATGCTTGAGAGATMGACAAAGG	382
OY	409	gaggacaggaanaaatgtatglccagatlaccogccccaacocacagaccaaaagt	468
Db	383	AAAACCAAGAAACAGCTGTATGCGAGATCCCGAACCCACACAGCTCGCGAGTA	442
OY	469	gaatttttgggaaacagctggatttgcagactgtgatacccggggttttgcacctga	528
Db	443	GAGTTCCTGGGGACCGCGGGTTTTGCGAGACTCTGGAATCCCGGATTTGCCAC	502
OY	529	gccccaactctaaccgctlaaccaaagaanaagggatctctcttggcctctgagcac	588
Db	503	GCTCCCTGTATCCACTAACCAAAGAGAGGGGAATTCACCT- GGATAGAGAAACATCAG	561
OY	589	aagcatcttgatgcatacaaaaaggccctgtctgtagcgcaactgcctctgscctcc	648
Db	562	CTAGCCTTTTGAACCTCTCAAAAAGGCACTGCTGCGAGGCTCGGCAATGGAG	621
OY	649	gtaactaaaccttaccccttatctatgtgatbagcgtlaaggagataagcccgagatt	708
Db	622	TTAAACAACCTTTACCCCTATACCTTATGATAAATAATGGAGTGGCAAGGGAG	681
OY	709	aaccaaaccttagaacaatlgagagagacttgttctcactcgtlcaagaagatgatcc	768
Db	682	ACCCAGGTTTGGGACCATGTGAAGGCGCCGGTAGCCTACTATATAAAGAA-- CTGACGCT	739
OY	769	gtagcagatgtgtgac 784	
Db	740	GTGGCCAGTGAATGGC 755	

RESULT	9
BH083773	
LOCUS	BH083773 653 bp DNA GSS 18-JUN-2001
DEFINITION	RPCI-24-293G13.TV RPCI-24 Mus musculus genomic clone RPCI-24-293G13 DNA sequence.
ACCESSION	BH083773
VERSION	BH083773.1 GI:14903370
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eumariota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 653)
AUTHORS	Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akintel,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Georgsgoris,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE	Mouse BAC End Sequences from Library RPCI-24
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPCI-24-293G13.TU

(pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.choi.org/bacpac/end_intro.html. Plate: 293 Row: 6 Column: 13 Seq primer: 17 Class: BAC ends.

FEATURES

Location/Qualifiers
1. 653

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-293G13"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pPRABAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pPRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 167 a 156 c 181 g 149 t

ORIGIN

Query Match 10.1%; Score 336.8; DB 13; Length 653;
Best Local Similarity 69.8%; Pred. No. 6e-57;
Matches 455; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

```

OY 280 ggtacgaagcactactgtcgtgaatgtctgctacgaagcctctgtaagaag 339
Db 2 ggaactgagagcctctgctgacgaactgagcttggggatgagcttggggcctaaag 61
OY 340 gcccaagattgacgagagagagtaacactggtgacagtttggggcggcagcaga 399
Db 62 gcacaaatttttcaaacagagagtaacctgacccggggtatgacccctccgaggggtaaaaga 121
OY 400 tggctgacgagagcagcagaagaacatgagtcacagatgacggcccaacacagcga 459
Db 122 tggctgacgagagcagcagaagaacatgagtcacagatgacggcccaacacagcga 181
OY 460 caagtgaagagagttttggggagcagctgagatttgcagactgtgacccgggtttgcg 519
Db 182 caggtgagcagagttttggggagcagctgagatttgcagactgtgacccgggtttgcg 241
OY 520 acctgagcagcagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 579
Db 242 atcttaccgagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 301
OY 580 gaggcaccagaagcattgtatgctacaagaagcctcctgagcagcactgctcctgccc 639
Db 302 gaggcaccagaagcattgtatgctacaagaagcctcctgagcagcactgctcctgccc 361
OY 640 ctccctgagcgttaacaccccttaaccttaltgtatgagcgtlaaggagagtagccga 699
Db 362 ttaccacagacttgaacttaaccttaltgtatgagcgtlaaggagagtagccga 421
OY 700 gggagtttaacccaacccctgagacatgagagagcagcgttgcctcactcgttaagaag 759
Db 422 gggagtttgaacccaacccctgagacatgagagagcagcgttgcctcactcgttaagaag 481
OY 760 ctgtatcctgtagcagcagcgttgcctgagcgttgcctgagcagcgttgcctgagcagc 819
Db 482 ttgagatccgctgtagtgatgagcagcagcgttgcctgagcagcgttgcctgagcagc 541
OY 820 ctgtcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 879
Db 542 ctgtcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 601
OY 880 gcatgagagacatcgttgcgagcagcagcagcagcagcagcagcagcagcagcagcagc 931
Db 602 gccctggaagatgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653

```

RESULT 10

AZ827524

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std. Error: 0.00
Plate: 0104 Row: E Column: 02
Seq primer: CGTTGTAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 700.
Location/Qualifiers
1. 700
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC2M0104E02"
/clone_lib="Mouse 10kb plasmid UUC2M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (9114/32114[9b]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

FEATURES

source

BASE COUNT

178 a 182 c 175 g 165 t

ORIGIN

Query Match 10.1%; Score 335; DB 13; Length 700;
Best Local Similarity 69.5%; Pred. No. 1.3e-56;
Matches 455; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

```

OY 460 caagtgaagagcttttggagcagcgtgagatttgcagcagcgtgagcagcagcagcagc 519
Db 8 caggtgagcagcgttgggagcagcgtgagatttgcagcagcgtgagcagcagcagcagc 67

```

[illegible]

RESULT	11
AZ897994/c	
LOCUS	
DEFINITION	AZ897994 689 bp DNA GSS 05-MAR-2001 RPCT-24-173A19.TV RPCT-24 Mus musculus genomic clone RPCT-24-173A19 DNA sequence.
ACCESSION	AZ897994
VERSION	AZ897994.1 GI:13216939
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus. 1 (bases 1 to 689)
AUTHORS	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akiret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shahtsbydn,A., Gebregorgrgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE	Mouse BAC End Sequences from Library RPCT-24
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPCT-24-173A19.TJ

Other_GSSs:RPC1-24-173A19.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igr.org
Clones are derived from the mouse BAC library RPC1-24. For BAC library availability, please contact Pieter de Jong (pdejong@email.cho.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tligr.org/cdb/BAC_endos/mouse/BAC_end_intro.html
Plate: 173 row: A column: 19

FEATURES	SOURCE	Seq primer: T7	Class: BAC ends.
Location/Qualifiers	1. .689		
/organism="Mus musculus"			
/strain="C57BL/6J"			
/db_xref="taxon:10090"			
/clone="RPC1-24-173A19"			
/clone_1lb="RPC1-24"			
/sex="Male"			
/cell_type="Spleen/Brain"			
/note="Vector: pTRABAC1; Site_1: BamH1; Site_2: BamH1; RPC1-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."			
BASE COUNT	168 a	174 c	170 g
ORIGIN			177 t

Query Match	10.0%	Score 333.2	DB 13	Length 689
Best Local Similarity	68.9%	Pred. No. 3.1e-56		
Matches 474	Conservative	0	Mismatches 208	Indels 6
QY 1933	gtgactcactagtgtaaagccgcgtcaataatcgggaacaaatctatctgtgttttglya	19922		
Db 688	GTGATTTTACTTAAGTTAACTGGAATGTATGTGAACAATATCTGTATGTATTTGTA	629		
QY 1993	gaacacfttcagatgtygtagaagctctacctaagaagaagacttcaacgtygtg	20523		
Db 628	GACACCTTTTATGAGATGGGCGCAGCATTTTCCACTAAACAGAACTGCCAGATTGTG	569		
QY 2053	gctaagaanaactctgaagaagaattttccaagaatttggataactcaaaagtiaaiaaglyca	21122		
Db 568	GCCAAAGAGATCCTTTGAAAGAAATCCGTCCCAAGATTGTGGATCCCTTAAGSTAACTCGGCTTC	509		
QY 2113	gacaatgctcagctcttcgttcgcaagtaagtcaagagacttggccaagaatltgsgyatt	21727		
Db 508	GACATATGACCAAGCCTTTGTGTGCCAGAGTAAAGTCAGGCGTTGGCACCTCAGTTGGGCATC	449		
QY 2173	gattggaactctgatttctgtatatacagaccccaaaagctcagaagaagtagagaatgaat	22333		
Db 448	GATTGGAAATTTACACTGTGCTTACCGCCCTCAAAAGCTCAGGACAGGTATAGAGGATGAAT	389		
QY 2233	agaacctaataagagagcccttaccaaattgacacagacagactggcataatgattgagtg	22922		
Db 388	AGGACCTTTAAAGAGACCTTTGACTTAATTAATGACATTTGAACCGGCGGAAACACTGGGTG	329		
QY 2293	gctctcctgccccttctgtccttttlaaggttlaagaaaccccttgaacagtttggctlyaac	23552		
Db 328	GCTTCTCCTCCTCTTGGGCTTTCGCCAGCCGGSAACACATCTGTGGAGGTTTGGGGCTCACT	269		
QY 2353	cccatatgaattgctctcaggggagaccccccgcttgcaggaattg-----ccttga	24002		
Db 268	CCTTTGAAGTTCGTATGAGAGACCTCCCCCTTATATGAAACCTGTAGAACATTATGTT	209		
QY 2407	catatgtctatgtatgcgtcgtcttccagccttcttctctlaagtgctcaagcgcgtcagtg	24662		
Db 208	TCCGACTCTGTGACCCTGTACTACCTCTCTTTTGTCTATTTCATTAATAAGGCCCTCAGAAGTG	149		
QY 2467	gtgaggcagcagcagcgttgaagacagctccggagagcctactacagagagagacttgcgaatt	25222		
Db 148	ATTAGAAACCCGATTTTGGGAAACATCTAAAGCAGCCCTATATCCCAAGGAGCACCGCAGTA	89		
QY 2527	ccacatgcctccaagtttgaagatcaagttcctatglttlaagcagccagtcaggaaaacc	25882		
Db 88	CCCCACGGGTTCCCAAGTTGGAGACAAAGTCTTGCTAGGGCGATAGAAACGGCAGCCTT	29		
QY 2587	gagactcgttgaagagacactatctcg 2614			
Db 28	GAGCCACGGTGGAAAGGACCTTATTTGG 1			

RESULT	12
LOCUS	A2717731
DEFINITION	A2717731 817 bp DNA GSS 24-JAN-2001
ACCESSION	RPCT-24-15511.1 TV RPCT-24 Mus musculus genomic clone RPCT-24-15511, DNA sequence.
VERSION	A2717731
KEYWORDS	A2717731.1 GI:12456936
SOURCE	GSS.
ORGANISM	house mouse, Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus 1 (Passes 1 to 817)
AUTHORS	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akincit,B., Levins,M., Tsengye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,I., de Jong,P. and Fraser,C.N.
TITLE	Mouse PAC End Sequences from Library RPCT-24
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPCT-24-15511_TV

Query Match	Best Local Similarity	Matches	501;	Conservative	0;	Mismatches	261;	Indels	3;	Gaps	1;
QY	809	ctgtggccatctgtgtcaagagcgtgtgcaaaattgactttgaggagaaatataactgttaa	868	1	CAGTAGCCCTGTGTTAAAGATGCTGCACAATTTGCATGTGGACACAGTGCACGTG	60					
QY	869	taggcccccatgcatctggagaacatcgtttgagcaagccccagacccgatgatgatacgaag	928	61	TGGCCCCCTATGCTTTAGAAAGTATGCGCGCTGCACCCACCTGCACAGATGGATGACAAATG	120					
QY	929	cccgcatgacccacatcataaagccctgctttctacagagagaggttactcttgctccacag	988	121	CCCGAATGCCACACTATATCAGAGCCCTGCGCTTAAAGACGCTTTAACTTTGGCCCCCTG	180					
QY	989	ccgctctaaacctgtccactctcttgctctggaagagactgaaacagtgactcatgtat	1048	181	CCATCTCAACCCACACTTACCTTTCCCTCTTAACAAAGATGATTCGTCACAGATCATAT	240					
QY	1049	gccatcaactattgtttgagagagactgggtgcggaagagaccttaagacataccgctga	1108								

Db 241 GATATGACATCCCTCGGTGAAGAACTGGGAGCCAGAAAGTATCTGACTGATACCAACCTGCG 300

Qy 1109 ctgagaaatgctaactgttacttgactgaagaaagactatgltgltgaagttaagaga 1168

Db 301 CTGG---ACCTCCAGTTGGTACAGGAGACGGACAGTTTCTGTATTAAGGGAAGCGAA 357

Qy 1169 tggctggggcggcagttgttgaacgggagcccgacagatctgggcccagcagccttgcggag 1228

Db 356 AGCGTGGAGCTCGGGGTGTGACTGTGGGAAAAGGTATTTGGGCAAGCCCTTTGGCTCGAAG 417

Qy 1229 gaacttcagcgcgaagaagctgagctcatalgagccctcaagcaagccttggcgcttcggag 1288

Db 418 GAACGTGGGCACAAAGGCTGACTTAAAGCGTTATTCAAGCCCTCCGAAGCGTTAAG 477

Qy 1289 ggaatcctaataactttatatacgcagcagtgatgctcttgcagctgcacagtlacag 1348

Db 478 GTTAACTATTTAACCTTACACTGCACGCCCTATGCTTTGGCTTACACACATCATCATG 537

Qy 1349 gggccatcatalaanaaagggggtgtgcttactcaagcagggagggaaataagaacaag 1408

Db 538 GGGGCATCTTACAGGCAGCAAGGGCATTTGAGCTGGGCGTGAAGACATTAAAAACAAG 597

Qy 1409 aggaattctaagctattatagaagccttaccttggccaaaaagcgtagtattatacct 1468

Db 598 AAGAAATTCTGGCCCTGTTGGAAACCATACATGCACTTAAACAAAGTGAAGCCATCATTCAC 657

Qy 1469 gtccctgagacatcagaagcccaagatctcatalactatagaaggaaacagatgtgcagg 1528

Db 658 GCCCGGGCCACCAAGAGGAGAAAGACTTGGTGGCCAGGGGGAACCGAATGGCAACATCAG 717

Qy 1529 ttgcgaagcagcagcccgagctgttacctctgcctcataatag 1573

Db 718 TAGCAAAACAAGTGTCTCAAGGGCCCATGATCTTAACCTGAAGAAAG 762

RESULT	13			
A2769626/c				
LOCUS	A2769626	692 bp	DNA	16-FEB-2001
DEFINITION	M05700.04R mouse 10kb plasmid UNGCM1 library Mus musculus genomic clone UNGCM05700.04 R, DNA sequence.			
ACCESSION	A2769626			
VERSION	A2769626.1	GI:12889953		
KEYWORDS	GSs.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 652)			
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinley,A., von Niederhausern,A. and Wright,D., Weiss, R.			
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert length: 10000 Std Error: 0.00 Plate: 0570 row: J column: 04 Seq primer: CACACAGCAAAACACGCTATGACC Class: plasmid ends High quality sequence stop: 692. location/Qualifiers 1. 692 /organism="Mus musculus" /strain="C57BL/6J"			
FEATURES				
SOURCE				


```
/organism="Mus musculus"  
/strain="C57BL/6J"
```

419 CCCACACATCGCTCCAGGTAGATTCGTTCTACATTAGAGCTCTCCACGCCGAAAAAC 478

Search completed: February 23, 2002, 23:36:01
Job time: 9362 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:36:01 ; Search time 6550.43 Seconds
(without alignments)
13466.612 Million cell updates/sec

Title: US-09-171-553B-3

Perfect score: 8209

Sequence: 1 gtggtgacgactgtggcc.....aaaaaaaaaaaaaaaaaaaa 8209

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estfun:*

2: em_esthum:*

3: em_estin:*

4: em_estom:*

5: em_estpl:*

6: em_estba:*

7: em_estro:*

8: em_estov:*

9: em_hic:*

10: gb_est1:*

11: gb_est2:*

12: gb_hic:*

13: gb_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_rod:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	600	7.3	2085	12 AK014027	AK014027 Mus muscu
2	594.6	7.2	857	11 B1184013	B1184013 UNL-P-FN
3	552.4	6.7	2096	12 AK011989	AK011989 Mus muscu
4	473.6	5.8	549	10 AW657531	AW657531 110922 MA
5	472	5.7	2935	12 AK004927	AK004927 Mus muscu
6	420.6	5.1	467	11 BF441465	BF441465 257856 MA
7	419.8	5.1	466	11 BF441468	BF441468 257862 MA
8	419.4	5.1	465	11 B1399234	B1399234 MI-P-AV1-
9	404.2	4.9	801	13 BH063533	BH063533 RPCI-24-3
10	401.6	4.9	452	11 BF441466	BF441466 257857 MA
11	385	4.7	576	13 AZ730616	AZ730616 RPCI-24-1
12	383.6	4.7	564	11 B1182930	B1182930 UNL-P-FN-

13	380.8	4.6	864	13 BH056750	BH056750
14	380.4	4.6	729	13 AZ318021	AZ318021
15	373.2	4.5	673	13 AZ329067	AZ329067
16	371	4.5	447	11 B1401063	B1401063
17	367	4.5	679	13 B1121259	B1121259
18	366.6	4.5	646	13 AZ967275	AZ967275
19	358.6	4.4	602	11 BF712162	BF712162
20	355.4	4.3	564	11 B1181099	B1181099
21	355.4	4.3	685	13 AZ636636	AZ636636
22	354.8	4.3	564	11 B1186129	B1186129
23	352.4	4.3	644	13 B1124013	B1124013
24	352	4.3	376	11 BF441469	BF441469
25	351.4	4.3	690	13 AZ627566	AZ627566
26	349	4.3	799	11 B1107577	B1107577
27	348.8	4.2	705	13 AZ337037	AZ337037
28	343.8	4.2	598	13 AZ959941	AZ959941
29	343.2	4.2	791	13 BH090616	BH090616
30	341.6	4.2	842	13 B1106844	B1106844
31	340.2	4.1	789	13 AZ715956	AZ715956
32	340.2	4.1	975	11 BF784826	BF784826
33	338.8	4.1	755	13 AZ715272	AZ715272
34	338.6	4.1	601	13 AZ650638	AZ650638
35	338	4.1	617	13 AZ420279	AZ420279
36	336.8	4.1	653	13 BH083773	BH083773
37	336.6	4.1	632	10 AL134888	AL134888
38	336.6	4.1	657	13 BH096361	BH096361
39	336.4	4.1	813	11 B1152889	B1152889
40	336.2	4.1	682	13 AZ980516	AZ980516
41	335	4.1	700	13 AZ827524	AZ827524
42	333.2	4.1	689	13 AZ897994	AZ897994
43	331.4	4.0	817	13 AZ717731	AZ717731
44	330.8	4.0	692	13 AZ769626	AZ769626
45	330	4.0	688	13 AZ912534	AZ912534

ALIGNMENTS

RESULT 1

AK014027	AK014027	2085 bp	mRNA	HTC	05-JUL-2001
LOCUS	Mus musculus	13 days embryo head cDNA, RIKEN full-length enriched library, clone:3110009E22, full insert sequence.			
DEFINITION	AK014027				
ACCESSION	AK014027.1	GI:12851651			
VERSION					
KEYWORDS	CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) 13 days embryo head cDNA to mRNA, clone:3110009E22.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	1 (bases 1 to 2085)				
JOURNAL	High-efficiency full-length cDNA cloning				
MEDLINE	Methods in enzymology. 303, 19-44 (1999)				
PUBMED	99279253				
REFERENCE	2 (bases 1 to 2085)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3 (bases 1 to 2085)				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,				

JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS	Genome research. 10 (10), 1617-1630 (2000) 20499374 11042159 3 (bases 1 to 2096) Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Ishiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Akiyama,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Todawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome research. 10 (11), 1757-1771 (2000) 20530913 11076861 4 (bases 1 to 2096) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5 (bases 1 to 2096) Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Harada,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sugabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ , Tel:81-45-503-9222, Fax:81-45-503-9216) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. Location/Qualifiers 1..2096 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /db_xref="MGD:MGI:1905349" /db_xref="MGD:MGI:191970" /clone="2610305J24" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 days embryo" 688..>2096 /note="putative" /codon_start=1 /protein_id="BAB27961.1" /db_xref="GI:12848455" /translation="MGQIVSTPLSTKDHWTVDVRGQNLVSKVKKPKMWTCTSEWPFVFGWGAEGFYFLPTIRAKIAIVFEGSGHPDQOQIMWEDLIARPPPPVRFPLPLRPTCKTALIQENGEKRPKPIIGRDDDSPTVPKPKIYSEIEEPPWPDTPQPPVAPQPSAPSGPLPQAPAGGGSPGTGTSRRVGTPEGPADSWPLRAIGAPADPNSLPQLQWPPSSSDPLNNKAAHHPPFSENPAGLTGLVLSLMYSHOPTWDDCOQLQTLFTTEPRILLKARNVRDEAGRPVQTAEIDEGPLTPPMDYWTASGRRLSNVRRVLVLAGRAAQPTNLAKRVEMQGAETPPSVFLEMEAYRRYPTFPDPTSEQRASVIMAF7GQAPDRKLRQLEQIGQETIIRDVREAEKVVHRRDETCKLEREKKREKREEDRRDRGEKVLTRILAAVG"
JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS	BASE COUNT 500 a 527 c 577 g 492 t ORIGIN
JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS	Query Match 6.7%; Score 552.4; DB 12; Length 2096; Best Local Similarity 63.5%; Pred. No. 1.1e-101;

Query Match 6.7%; Score 552.4; DB 12; Length 2096;
Best Local Similarity 63.5%; Pred. No. 1.1e-101;

QY 1299 ctacacgggggttggtgagtccttattgttctctacacagcctactctggtgatgtattctaa 1358
Db 1 CTCACGGCGTGTAGTGGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGATTCAG 60
QY 1359 cagctctcagacactcttcaaacacggagagcagagagattctgttagagctaga 1418
Db 61 CAGCTGCTGAGACACTCTTCAACCCGAGGAGCGAGAGAATCTGTAGAGGCTAGA 120
QY 1419 aaaaaatttctctgggcccagcggcagccacccagcttgcaaaatgagattgacatggga 1478
Db 121 AAAAAATGTTCTTGGGCGGACGGCGACCCACGCTGTCAAAATGAGATTGACATGGGA 180
QY 1479 ttctccttgactgcgcgcgggttggtgactacaacacggtgaaagtgaggagccttga 1538
Db 181 TTTCCCTTGACTGCGCCGAGTGGGACTACAAACACGCGGAAAGGTAGGAGAGCTTGA 240
QY 1539 atctatgcagactctgttgccggtctccgggcccctcaagacgcccactaattg 1598
Db 241 ATCTATCGCAGGCTCTGCTGCGGGTCTCCGAGGCGCTCAAGACGGCCCTACTAATTG 300
QY 1599 gctaaagtaagagaggttgatgcaggagcagcaaacctccctcggtattcttctgagag 1658
Db 301 GCTAAGCTAGAGAAATGATGACGAGGACTGAATGAACCCCTCTGTTTCTTGAGAG 360
QY 1659 ctcatggaacccctcagggggttccacccctttgtactactcagagccagcaagaagcc 1718
Db 361 CTCTTGAAGCCTTCAGGCGGTACACCCCTTTTGTATCCACCTCAGAGGCGCTAAAA 420
QY 1719 tcaatggccttcctcattggcaggtcaggtcgtcgtcgtgatatca 1761
Db 421 TCAGTGGCTTTTGGCTTTATAGACAGCTCAGCCTTGGATATTA 463

RESULT 8
BI399234/c

LOCUS BI399234 465 bp mRNA EST 14-AUG-2001
DEFINITION MI-P-A11-nrp-a-02-0-UI.s1 MI-P-A11 Sus scrofa cDNA clone
MI-P-A11-nrp-a-02-0-UI 3', mRNA sequence.
ACCESSION BI399234
VERSION BI399234.1 GI:15178301
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 465)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildeer Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: ktuggle@iastate.edu
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
University of Iowa Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.

FEATURES
source

Location/Qualifiers
1..465
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-A11-nrp-a-02-0-UI"
/clone_lib="MI-P-A11"
/lab_host="DH10B (Life Technologies)"

/note="Vector: pMT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-A11
library is normalized library derived from the MI-P-A10
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
http://pigest.genome.iastate.edu/. The procedure used to
create this library has been previously described (Bonaldo
, Lennon and Soares, Genome Research 6: 791-806, 1996)
TAG_SEQ=None found"

BASE COUNT 137 a 122 c 100 g 106 t
ORIGIN

Query Match 5.1%; Score 419.4; DB 11; Length 465;
Best Local Similarity 95.4%; Pred. NO. 1.le-74;
Matches 432; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 493 gcttggaaagcggcggcgcgtgctgctgagatctgttggttctctcgtgtg 552
Db 465 GCTTGGAAAGACGTCGACGGCTCAGTGTCTGTGGATCTGTGTGTTTGTGTG 406
QY 553 tcttggcttggctccttctctacagttttaataatggacagacagactaccccc 612
Db 405 TTTTGTCTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 346
QY 613 ttagtctgactctgacacattgactgaagttagatcagggcctataattgtcagttc 672
Db 345 TTAGTTTGAATCTGACCATTCGACCATTCGACCATTCGACCATTCGACCATTC 286
QY 673 aggttaagaagggaccttggcagacttctgtgcctcctgaatggcacaattcagttg 732
Db 285 AGGTTAAGAAGGACCTTGGCAGACTTCTGTCTGTCTGTCTGTCTGTCTGTCTG 226
QY 733 gatggccatcagaggggacctttaattctgaaattatctggtcttgaagcaattcatt 792
Db 225 GATGGCCATTCAGAGGGGACCTTTAAATCTTGATGATTATCTGCTGCTGTAAAGCAATATT 166
QY 793 ttcagactggacccggcctctcctcctgactcagagacccctatctctacgtggcagatt 852
Db 165 TTCAGACTGGACCCGGCTCTCATCCGATCAGGACCCCTATATCTTACGTGGCAAGATT 106
QY 853 tggcagaagatcctccgcctgggttaaacatggcctaaataaaccaagaaagcaggtc 912
Db 105 TGGCAGAGGATCTCCGCCATCGGTTAAACCATGCTGTAATAAGCAAGAAAGCCAGGTC 46
QY 913 ccggaatcctggctcttggagagaaaaaacaac 945
Db 45 CCCGAATCTGCTCTTGGAGAGAAAAAACAAC 13

RESULT 9

BI063533
LOCUS BI063533 801 bp DNA 18-JUL-2001
DEFINITION RPCI-24-370H8.TJ RPCI-24 Mus musculus genomic clone RPCI-24-370H8,
DNA sequence.
ACCESSION BI063533
VERSION BI063533.1 GI:14878283
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 801)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-370H8.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research

Db 61 CAGCTGCTGCAGACACTCTTACAAACGAGGAGCGAGAGAAATCTGTTAGAGGCTAG- 119
 QY 1419 aaaaatttctgggcccagggcgacccacagctgtgcaaatgagattgacatggga 1478
 Db 120 AAAAAATGTTCTGGGGCCGACGCGGCGACCCACACAGTTGCAAAATGAGATTGACATGGGA 179
 QY 1479 ttcccttgactgcgcgcggttgaggactacaacacgctgaagtgaggagagcttgaaa 1538
 Db 180 TTTCCCTTGACTCGCCCGGATGGGACTACAACACGCGCCGAAGGTAGGAGAGACTTGAAA 239
 QY 1539 atctatgcagcagctctggtggtggtctccgggggctcctcaagagcggccacataattg 1598
 Db 240 ATCTATGCCAGGCTCTGTTGGGGGCTCTCCGAGGGGCTCTCAAGACGGCCCACTAATTG 299
 QY 1599 qctaaggttaagagagtgatgcagggacgaacgaacacccctccctcggtattcttgagag 1658
 Db 300 GCTAAGGTAAAGAAATGATGACAGGAGCTGAATGAACCCGCCCTCTGTTTCTTGAGAG 359
 QY 1659 ctcataagacccctcagggcgtttcaccccttttgatctcctacccagagcccgagaagcc 1718
 Db 360 CTCTTGAAGCCCTCAGGCGGTACACCCCTTTTGATCCACCTCAGAGGCGCTAAAAAGCC 419
 QY 1719 tcaatggcctggccttcattggcagctggc 1750
 Db 420 TCAGTGGCTTTGGCCCTTATAGGACAGTCAGC 451

RESULT 11

AZ730616 576 bp DNA GSS 25-JAN-2001
 LOCUS RPCI-24-149D1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-149D1,
 DEFINITION DNA sequence.
 AZ730616
 ACCESSION AZ730616.1 GI:12491859
 VERSION
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 576)
 AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-24-149D1.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 149 row: D column: 1
 Seq primer: Sp6
 Class: BAC ends.

FEATURES
source

Location/Qualifiers
 1. 576
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-149D1"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA.
 BASE COUNT 161 a 139 c 121 g 155 t
 ORIGIN

Query Match 4.7%; Score 385; DB 13; Length 576;
 Best Local Similarity 82.5%; Pred. No. 1e-67;
 Matches 453; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

QY 5040 ggaataacctaaaggttaggggttcagacaactgtcagcttctccggttcctccaggttaagtcag 5099
 Db 1 GCATACCCCAAGTAAATCGGGTCCATTAATGACCTGCTTTGTTGGCCAAAGTAAGTCAG 60
 QY 5100 ggactggccaaagattgggggattgattggaactgcattgtgcatacagaccccaagc 5159
 Db 61 GGACTAGCCACAGATATTGGGGATTGATTAGAAGTTACATTGTGTATACAGGCTCCCAAC 120
 QY 5160 tcaggacaggttagagagatgaatagaaccattaaagagacccttaacaaattgaccaca 5219
 Db 121 TCAGGACAGGTAGAGAGGATGAATAAACCATAAAGAGACCCCTTACCAAAATTG-CTGCT 179
 QY 5220 gagactggcattaatgattggatggctctcctgccttctgctttttaggtgaggaac 5279
 Db 180 ATGACTGCCACTAATGATTGGATAGCTCTCTCCCTTTGGCTCTTCAGGGTTAGAACC 239
 QY 5280 accctggacagtttgggctgacccctatgaattgctctacgggggggggggggggggggg 5339
 Db 240 ACCCTGGACAGTTTAGACTGAACCCCTATGAATCACTCTATGGGGGTCTCTCCATTTG 299
 QY 5340 gcagaaattgctttgcacatagctgctgctgctgtcttccagcctttgtctctagg 5399
 Db 300 GTAAAAATAGCCCTCTACATAGTCTGCTGCTGCTTTCCAGCCCTTTCTCTAGT 359
 QY 5400 ctcaaggcgtctgagtgagtcagcagcagcgtgggaagcagctccggagggcctactca 5459
 Db 360 CTCAGGTCCTCGAATGGTGAGACAGCAAGTATAGAGACAGCTCCAGAAATGCCACTCA 419
 QY 5460 ggagagagacttgcagttccacatcgttcccaagttggagattcagttatgttagagc 5519
 Db 420 AGAGAAGACCTGGGAAGCCCACTCTCTTCCAAAGTTATAGATTTCAGTCTACATTAGAGC 479
 QY 5520 caccgtgcagaaacctgcagacacgcggtgggaaggacctatctcgtactttgaccaca 5579
 Db 480 CTCACACAGAAAACCTTTAAGACTCAGCGGAAAGCCCTTATCTGTGAATCTTTACCACC 539
 QY 5580 ccaacgct 5588
 Db 540 CTAATTCTCT 548

RESULT 12

BI182930/c 564 bp mRNA EST 10-JUL-2001
 LOCUS UNL-P-FN-br-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 DEFINITION UNL-P-FN-br-h-08-0-UNL 3', mRNA sequence.
 ACCESSION BI182930
 VERSION BI182930.1 GI:14657339
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 564)
 AUTHORS Caetano,A.R., Johnson,R.K. and Pomp,D.
 TITLE Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA

Tel: 402 472 6416

Fax: 402 472 6362

Email: dpoem@unl.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized porcine ovarian follicles library

Seq primer: M13 -29

POLYA-yes.

FEATURES

Location/Qualifiers

1. .564
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection Lines"
 /db.xref="taxon:9823"
 /clone="UNL-P-FN-br-h-08-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
 TAG_LIB=UNL-P-FN
 TAG_TISSUE=porcine ovarian follicles
 TAG_SEQ=CACACT"
 BASE COUNT 141 a 129 c 143 g 151 t
 ORIGIN

Query Match

Best Local Similarity 4.7%; Score 383.6; DB 11; Length 564;

Matches 477; Conservative 0; Mismatches 9; Indels 78; Gaps 1;

QY 7720 cctgctgctgcagtaaataggtagaaggtcacacttcctattgttcacgggctgcta 7779
 |||||
 DB 564 cctgctgctgcagtaaataggtagaaggtcacacttcctattgttcacgggctgcta 505
 QY 7780 tcttgctgaagataacaggaaatgagttgactaattgcttatctgagttctgtaa 7839
 |||||
 DB 504 tcttgctgaagataacaggaaatgagttgactaattgcttatctgagttctgtaa 445
 QY 7840 aactgactggccaccatagaagaattgattacacattgacagccctagtgacctatctcaa 7899
 |||||
 DB 444 aactgactggccaccatagaagaattgattacacattgacagccctagtgacctatctcaa 385
 QY 7900 ctgcaatctgtaactctgccaggagccacgcagatgctggagacctccggagctatttaa 7959
 |||||
 DB 384 ctgcaatctgtaactctgccaggagccacgcagatgctggagacctccggagctatttaa 325
 QY 7960 aatgattgggtccacagagcgggctctcgatatttttaaatgattggtccatgagcgc 8019
 |||||
 DB 324 aatgattgggtccacagagcgggctctcgatatttttaaatgattggtccatgagcgc 265
 QY 8020 g-----
 DB 264 gggctctcgatatttttaaatgattggtccacagagcgggctctcgatatttttaaat 205
 QY 8021 -----ggctctcgatatttttaaatgattggtccacagagcgggctctcgatatttttaaat 8061
 DB 204 gattggtccacagagcgggctctcgatatttttaaatgattggtccacagagcgggctctcgatatttttaaat 145
 QY 8062 ctttatttgaaacccataaaagctgtcccgattccgcactcggggcgagctcctctac 8121
 |||||

Db 144 CTTTGTGTGAACCCCAATAAAAGCTGTCCGACTCGGCACACACGCCGCCGACTCTCTAC 85
 QY 8122 ccctgctggtgtacagctgtggccccagcgcgcttggaataaaatcctctgtgtt 8181
 |||||
 Db 84 CCCTGGTGGCTTACGACTGTGGCCGCCAGCGCTTGGAAATAAAATCCTCTGTGT 25
 QY 8182 tgcatacaaaaaaataaaaaa 8205
 |||||
 Db 24 TGCATCAAAAAAATAAAAAA 1

RESULT 13

BH056750 864 bp DNA GSS 18-JUL-2001
 LOCUS RPCI-24-94C16.TVB RPCI-24 Mus musculus genomic clone RPCI-24-94C16,
 DEFINITION DNA sequence.
 ACCESSION BH056750
 VERSION BH056750.1 GI:14864050
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 864)
 REFERENCE Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 AUTHORS Mouse BAC End Sequences from Library RPCI-24
 TITLE Unpublished (1999)
 JOURNAL Other_GSSs: RPCI-24-94C16.TJ RPCI-24-94C16.TV
 COMMENT Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 94 row; C column; 16
 Seq primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1. .864
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-94C16"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

FEATURES

BASE COUNT 237 a 221 c 209 g 197 t
 ORIGIN
 Query Match 4.6%; Score 380.8; DB 13; Length 864;
 Best Local Similarity 66.5%; Pred. No. 7e-67;
 Matches 577; Conservative 0; Mismatches 287; Indels 4; Gaps 2;
 QY 3414 gtgaagaagttttggagcagctgatttgcagatggtccgggttgagacc 3473
 |||||
 Db 1 GTAGCTGAGTTCTGGGAACGGGCGCTTTGTAGACTCTGAATACACGGGTCACAATG 60
 QY 3474 ttacagccccactctaccgcgtaaaccaagaaaggggattctctctgggctcctgag 3533
 |||||

```

Db      61 TTAGGGGACCTCTGTACCCACTCACTAAGAAAAAGTCCGCTACCTGACAGACAGA 120
Qy      3534 gaccgaagcattgatgatacaaaagccctcgagcgcctcgtcgtcgcctc 3593
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 CACCAAAAAGACTTTAACAAGCACTAAGGCTGCTGCTGCGAGCCAGCTTTGGCCCTG 180
Qy      3594 cctgacgtaactaaaccttacccttattatgatagagcgttaagagagatgacccgagga 3653
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      181 CCATATTTGACTAAGCCTTTACCATATACGTTGATGAGCCAGCTGGAGTACCCGCGAGA 240
Qy      3654 gtttaaccccaaaccttagagacatgagagagacccgtgttaccatgltcaagaagct 3713
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      241 GTCCTGACTCAAGACTCTG6GGCGCGAGAGAGGCTAGTGGCTACTGCTTAATAAAAAATG 300
Qy      3714 gatcctgagccagtgagtgccggtatgltgtaaggtatcgcagcgtgtgtccatagt 3773
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      301 GACCCAGTGGCCAGTGAATGGGCTCTGCTGCTTAAGGCAATGCGTGAATGAGCCCTACTT 360
Qy      3774 gtcaagagcgtgtaacaaattgacttggagacagaataataactgtatagcccccagca 3833
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      361 GTTAAGACTAGTGGCAAGCTTACTGTAGTACAGCATGAGTGTATATACCTCCCATGCG 420
Qy      3834 ttggagacatcgttcgcagaccccacacacagatgatagacacagcccgatgacccac 3893
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      421 TTAAGAGTATTGTATGACACCCCTGACCATTTGAGTACAAATGCCCGATGACCCAC 480
Qy      3894 tatcaagcctgctctctacagagaggtacacttgcctccacagcgcgtcctaaccct 3953
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      481 TACCGAGACCTGTGCTTAATGACCGAGTGAATTTGGCTTACCTGCCATCCACATGCT 540
Qy      3954 gccactctctgctgtaagagatgatacagatgatacactgatacctcaactatg 4013
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      541 ACCACCTTCTCCCAAGACGATGACTCTACTGCGGTACACCAATGCTGATATCCTG 600
Qy      4014 attgagagagatgggtgctcgaagagacttacaacacacacgcgtgactgagaagtcta 4073
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      601 GATGAGAACTGATGATCGAAGAGATCTGACTGACCAAACTTGGCC--AGGTGTCCT 657
Qy      4074 acctgttactgacgagagacagctatgtgtgtaagatgagagatgagctgagggagca 4133
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      658 GACAGGTACACAGATGATGACAGCTTTGTGTGAGGAGAAATTAAGACGAGAGAGCA 717
Qy      4134 gtgtgtgacggagacccgacagatcgtggccagacagcgtcgcggaagaaacttcagcgaa 4193
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      718 GTGCTTGACAGAAAAATGATTAATTTGGCAAGTTAACTTCCAGAGAACTCCGCTCAG 777
Qy      4194 aaggtgagctcatagccctcctacagcagcttgcgtcgtcgcggaaggaatccataac 4253
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      778 AAGGCTGAGCTCTGACCGCTACTAAGACTCTGCGCTGG--CAAAAGCTAGGCAATCAT 836
Qy      4254 attatacgaacagcaggtatgacttgg 4281
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      837 ATTTAAGACTGACAGGCACTATGCTTTG 864

```

```

RESULT 14
LOCUS   AZ318021 729 bp DNA GSS 29-SEP-2000
DEFINITION
clone UUCG1M0036004 R, DNA sequence.
ACCESSION
AZ318021
VERSION
AZ318021.1 GI:10367383
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 729)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb

```

```

JOURNAL
COMMENT
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0036 row: 0 column: 04
Seq primer: CACACAGGAAGACGCTATGACC
Class: plasmid ends
High quality sequence stop: 729.
FEATURES
Location/Qualifiers
1..729
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0036004"
/clone_1lb="Mouse 10kb plasmid library"
/sex="Male"
/lab_host="E. coli strain XL10-gold, T1-resistant, F-"
/notes="Vector: PMD29v; Purified genomic DNA from M.
musculus C57Bl/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
chemically-competent E. coli XL10-gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT
196 a 193 c 182 g 158 t
ORIGIN
Query Match 4.6%; Score 380.4; DB 13; Length 729;
Best Local Similarity 73.4%; Pred. No. 8,5e-67;
Matches 486; Conservative 0; Mismatches 176; Indels 0; Gaps 0;
Qy 1215 cccctcagatattgccccttctctcgcagatcctaataatggaactaacatccc 1274
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 CCTTTACAAATACAGCGCTTTTCCTCTCTGATTAATATGGAATAACCAACCT 124
Qy 1275 ccttctcgagagatccccaacgcctcacgaggttggagttccctatgttctctac 1334
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 CCTTTTCAGAGAACCCCTCTGCGCTTACTGGGCTCTTGAGCACTTAATGTTCTCCAT 184
Qy 1335 cagcctacttggatgatgttcaacagcgtctgcagacacacttcccaacagagagca 1394
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185 CAAACCACCTGGGATGATTTGTGACGACGCTTTTGACAGCTTTTTCACACCAAGAANA 244
Qy 1395 gagaagaatcctgttagagcgtgtaaaaaatgtctcgtgggcccagagcgacccagcg 1454
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 245 GAAAGATCTCTGATGAGAGCGAGAAAAATGTTTATGAGAGAGACACACCCACTGCC 304
Qy 1455 ttgcaaatgagatlgagcagtgatctccctgacccgcccgggttgagacaaacg 1514
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 305 CTCCTCAACCTGCTGACACAGGCTTTCCCTTTGAACGCCCACTGGGACTACAAAC 364
Qy 1515 gctgaagtagagagagcttgaataatcctacccagcgtctgtgtgaggtctccggg 1574
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 365 GCGAAGGTAGGAGCGCCCTTGTCTATCGCGGAGCTTATGCGAGCTCTCAGAGA 424
 Qy 1575 gctcaagagagccactaatttgtaagtaagagagtgatgagagagagagagaa 1634
 Db 425 GCCGTAGACGGCCACCAATTTGGCTAAGTAAGAGAGGCTTGGCAGGGGAGACTGAA 484
 Qy 1635 cctccctggtattcttctgagagagctatgagagagagagagagagagagagag 1694
 Db 485 CCACCTCAGTCTTCTTGTAGCGCTCATGAGGAGCATATTAAGAGGTACACCCCTTTGAC 544
 Qy 1695 cctcctcag 1754
 Db 545 CCCCTGAGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 604
 Qy 1755 gatacag 1814
 Db 605 GACATTAAGAAAAGCTTCAAGAGCTGAGAGGCTTCAAGATCATATGCTTCAAGATT 664
 Qy 1815 gtgag 1874
 Db 665 GTAAAG 724
 Qy 1875 aa 1876
 Db 725 GA 726

RESULT 15
 A2329067/c 673 bp DNA GSS 29-SEP-2000
 LOCUS 1M0053G10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC1M0053G10 F, DNA sequence.
 ACCESSION A2329067
 VERSION A2329067.1 GI:10389414
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 673)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

TITLE Unpublished (2000)
 JOURNAL Contact: Robert B. Weiss
 COMMENT University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

FEATURES
 SOURCE Insert Length: 10000 Std Error: 0.00
 Plate: 0053 row: 6 column: 10
 Seq primer: CGTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
 Class: plasmid ends
 High quality sequence stop: 673.
 Location/Qualifiers
 1..673
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0053G10"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g11473211419b1AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance.

BASE COUNT 165 a 173 c 149 g 186 t
 ORIGIN

Query Match 4.5%; Score 373.2; DB 13; Length 673;
 Best Local Similarity 78.1%; Pred. No. 2.5e-65;
 Matches 520; Conservative 0; Mismatches 123; Indels 23; Gaps 5;
 Qy 4851 ctaagggagagccaccagagcgtcctcgtgagagtgacttcagagagagagagag 4910
 Db 672 CTCAGGAGATGGCTGGCTTGAAGCTTACCTGGAGGTAGAGATTTCACAAATCTAAGCGAGCT 613
 Qy 4911 aatacagagacacatcatctatgttcttgtaagacaccttcagagagagagagag 4970
 Db 612 AGATACGACACACAGTACTTATTAGTGTATTAGATACCTTTTCAAGGTGAGTACAGCT 553
 Qy 4971 tctcctactaagaagaagagagagagagagagagagagagagagagagagagag 5030
 Db 552 TTTCCTACAG 505
 Qy 5031 ccaagattggag 5090
 Db 504 CCGAGGTTTGTGTATCCCAAGGTAATCAGTCAATTAATGAGAGAGAGAGAGAGAGAG 445
 Qy 5091 gtaagtcag 5150
 Db 444 GTAAAGTCAAGGAG 385
 Qy 5151 ccccaagagtcag 5210
 Db 384 CCCCAAGGTCAG 326
 Qy 5211 ttgaccag 5270
 Db 325 TTGGCAGCTGAG 266
 Qy 5271 gtgag 5326
 Db 265 GTTAAAGACATCCCTGAACAGTTTGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 206
 Qy 5327 acccccccgttgcag 5386
 Db 205 GGGGCTCCCACTGTAAATTAATCCTCTGACTGAGAGAGAGAGAGAGAGAGAGAGAG 146
 Qy 5387 ttgtctctcag 5442
 Db 145 TTGCTTCTTCTAGAGCTCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 86
 Qy 5443 tccggag 5500
 Db 85 TCCAGAGAGGCTACTCAG 26
 Qy 5501 ttcaagt 5506
 Db 25 TTGGGT 20

Mon Feb 25 07:44:16 2002

us-09-171-553b-3.feb22std.rst

Page 15

Search completed: February 23, 2002, 23:37:03
Job time: 9424 sec

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

FEATURES Location/Qualifiers

source 1..2096

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/db_xref="MGI:1905349"

/db_xref="MGI:1919770"

/clone="2610305J24"

/clone.lib="Riken full-length enriched mouse cDNA library"

/dev_stage="10 days embryo"

688..>2096

/note="putative"

/codon_start=1

/protein_id="BAB27961.1"

/db_xref="GI:12848455"

/translation="MGQIVSTPLSLTKDHWIDYRANGQNLVKKKKPMWTECSSEMPVFGVGPABESTFLPTIRAKAIVFQEGPGSHPDQPYIMVEDLARYPPWVRPLPPLRPGKILAIQENGEKERKPLGRDDRSPTVTPKPKISEIEPPMPDTPQPPYAPQPSAPSGPLPOAPAGGSGSTGRSRGVEGADSTVALPLAIGAPADPNSLOPLQYMFSSDLYNMKAHPPSPENPAGITGVESLAWSHOPTMDCOILQTLFTTEERERILLERKNVDAGRPVQTPAIDRGLPTPRMDYNTASRELSNTRVIVAGRGAROPTNLAKVREVMQATEPPSVFLERLMATIRYTPPTDSEGQRAVYMAFICQSAPIDKKILQRIEGLQDYTRDVRVRAEKVYHRETEKLEREKREEEDRRRQKEVLRILIAAVG"

BASE COUNT 500 a 527 c 577 g 492 t

ORIGIN

alignment_scores: Quality: 1559.50 Length: 473
Ratio: 4.051 Gaps: 7
Percent Similarity: 81.395 Percent Identity: 65.328

alignment_block:
US-09-171-553B-4 x AK011989 ..

Align seg 1/1 to: AK011989 from: 1 to: 2096

1 MetGlylnThrValIthrProLeuSerLeuThrLeuAspHisTrpTh 17
|||||
688 ATGGAGACAGACCGTCTACTCTTATCTTGTGACTAAGACCACTTGAC 737
17 rGlnValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
|||||
738 GGAGCTTAGGGCTAGAGACAAATTTGTCAATAAGTGAAGAAAGC 787
34 rOTPGlnThrPheCysAlaSerGlnTrpProThrPheAspValGlyTrp 50
|||||
788 CATGATGAGACTTCTCTCTCAGAAATGGCTGTTTGGAGTAGGTGG 837
51 ProSerGlnGlyThrPheAsnSerGlnIleIleLeuAlaValLysAlaI 67
|||||
838 CCACCAAGAGAACTTTTACTTACCACCATTAAGGCTGTGAAGGCAT 887
67 eIlePheGlnThrGlyProGlySerHisProAspGlnGlyProTyrIle 84
|||||
888 TGTITTCAGGAAGACCAAGGTCGATCCAGACCAACACCGTACATCA 937
84 euthrTrpGlnAspLeuAlaGluAspProProThrValLysProTrp 100
|||||
938 TGGATGAGGAGACTTGGACCGCTACCCACCCCGTGGGTTCCGCCATTC 987
101 LeuAsnLysProArgLysProGlyLysProArgIleLeuAlaLeu..... 114
988 CTC...CCGCTCTCCGCTGCGACCAAGATTCAGCATCAAGAAAA 1034
115 .GlyGluLysAsnLys.....His 121
|||||
1035 TGGTGGAAAGAGAAACCGAACCCGCTCGGAGAGATGATGATGCA 1084
121 eValGluLysValGluProSerSerSerTyr.....LeuProArgAsp 135

|||||

1085 GCACACACAGTACCAAAACCCCAAGATCTATTAGATGAAAGAACCC 1134
136 ArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThrProle 152
|||||
1135 CCTGAGTGGCCGATACCCCTCAACCCCACTTATCTCCCAAGCCCA 1184
152 uSerSerThrGlyGlyScyGluGlyThrSerAlaProProGlyAlaPro 169
|||||
1185 ACCTTCAGCTCCCGGAGCCCTGCTCAGGCCCGCGGAGGA.... 1230
169 aValGluGlyProAlaGlyThrArgSerArgArgGlyAlaTrpPro 185
|||||
1231GGGGGCTCCACAGAAAGAACGCGGAGAGATCACCCCT 1275
186 Glu...ArgThrAspGluIleAlaIleLeuProLeuArgThrGly... 200
|||||
1276 GAGGGGCTCCGATCACTGCGGCTCCCTCAGGCTATTGGGGC 1325
201 .ProProMetProGlyGlyLeuGlnProLeuGlnTyrTrpProPhe 217
|||||
1326 TCCCGCTGGCCATCCAAATAGTCTACAGCCCTACAGTATGGCTTTT 1375
217 eSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
|||||
1376 CCTCTCTGACCTTCACTCAAGTGAAGCTATCACCCCTTTTCAGAA 1425
234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisG 250
|||||
1426 AACCTGCAGAGACTCACTGGGTGTTGAATCATTAATATTCACACCA 1475
250 nProThrTrpAspAspCysGlnGlnLeuGlnThrLeuPheThrThg 267
|||||
1476 GCGGACTGGGATGATCCAGCAGCTTCTGACAGCTATTCACACACCG 1525
267 LuGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAla 283
|||||
1526 AGGAGAGAGAGAGGATTCCTCGAGGCTCGGAAATGTCGAGACGAG 1575
284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProle 300
|||||
1576 GCTGGGCGCTGCTCCAACTCCAGCTGAGATGATGAAGATTTCCGCT 1625
300 uThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeu 317
|||||
1626 AACCCGCGCCGATGAGATTATATAGCCATCAGATGAGGAAACACTGT 1675
317 ySIIeTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
|||||
1676 CCAATTATATGCGGCGCTCAGTGGCGGCTCAGAGAGTCCCGCGAG 1725
334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGlu 350
|||||
1726 CCCACCAATCTGCCAAGGATTAAGAGAGGTTATGAGGAGCAGCTGACC 1775
350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThr 367
|||||
1776 CCCCTCAGCTTCTCTTGAAGGCTCATGAGGCTTATAGGAGATTAACC 1825
367 ropPheAspProThrSerGlnAlaGlnLysAlaSerValAlaLeuAlaPhe 383
|||||
1826 CATTGAGCCCACTGAGGCTCAAGAGGCTCAGTAAATTATGCGCTTC 1875
384 IleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 400
|||||
1876 ATGGGCGCTGCTGCTGACATTAGAGAGATTAAGCCCAATTAAGAGG 1925
400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysVal 417
|||||
1926 CTTGCAATTAACACCATTAAGGATGATGATGAGAGGACAGAAAGTGT 1975
417 yTyrArgArgGluThrGluGluGluGluGluGluGluGluGluGlu 433
|||||


```

556 TGTCTTTCAGATGGAAGCTGGTCCACCCAGACAGACACCCCTACATAA 605
84  eutThrTrpGlnAspLeuAlaGluAspProProProProValylsProTrp 100
   :::::::::::::::::::::::::::: |||||:::
606  CTGTGTGGAGAACCTCGTCGAAATTCACCCCTGATCAACACCGTGG 655
101  LeuAsnLysProArgLysProGlyProArgGlyLeuAlaLeuGlyGlu.. 116
   :::: ||| ||||| ||| ||||| ::::
656  ATCCCAAAA...CGGAAGCCAGGTTCTCGAGCCCTGGCTCCGACAGC 702
117  ....LysAsnLysHisSerAlaGlyValGlu..ProSerSerS 130
   ||| ||| ||||| ||||| |||||
703  AAAAGTGAACCAAGAGATCTCTGAAAAGAGCCTTGGGCGCT 752
130  eTyLeuProArg.....AspArgGlyAlaAlaAspLeuAla 142
   || |||||:::
753  CAGCTCTCCCAAGATCTATCCGAGATTGAAGGCCCTCCGAGTGGCA 802
143  GlyThrProThrCysSerProThrProLeuSerSerThrGlyCysGly 159
   ||| ||| |||
803  GAACCTCCTCTCCTATCCCTGCC...CC 831
159  uGlyThrSerAlaProPro.....GlyA 167
832  TCCAGCTTCACGCTCCACAGCTTCCAGAGCCGCCCTGGGCTGGT 881
167  LaProValValGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
   :::: ||||| ||||| ||||| ||||| ::::
882  GCGGACAGATTGGAGACCGGGCTGGACCAAGCCGAGGCCACAG 931
184  ThrProGluArg.....ThraspGlyLeuAlaLeuProLeuAr 197
   :::: ||| :::: |||||
932  ACTCTCTCGAGATGGGCTTCCAGACACATGCTGCTTCTTCTTG 981
197  gThrTyGlyProPrometProGlyGlyLeuGlnProLeuGlnTyrT 214
   :::: |||||
982  TGCAGTGGAGACCGGCGAGACCTCAGCTCTCCAGACTTCAAT 1031
214  rProProSerSerAlaAspLeuTyrAsnTrp.LysThrAsnHisProPr 230
   ||||| ||||| ||||| ||||| |||||
1032  GCCCATCTTCATCAGCTGATCTCTATTAACGTGAAT.....CC 1072
230  oPheSerGluAspProGlnArgLeuThrGlyLeuValGluSerLeuMetP 247
   ||||| ||||| ||||| ||||| |||||
1073  TTTTTCAGAAAATCCAGCTGACCTAACCAGCCTGGTGGAGTCATTAAT 1122
247  heSerHisGlnProThrTrpAspAspGlnGlnLeuGlnThrLeu 263
   ::||| ||||| ||||| ||||| ||||| |||||
1123  TATCTCATCAGCCACTTAGGATGATTCACAAGCTTTTGCATGTTCTT 1172
264  PheThrThrGluGluArgGlyLeuLeuGlnGlnAlaArgLysAsnVal 280
   ||||| ||||| ||||| ||||| ||||| |||||
1173  TTTACACAGAAAGAAAAGAAATCTCTTAAGGGCTCAGAGACGT 1222
280  LProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGlnLeuAspMetG 297
   ||||| ||||| ||||| ||||| ||||| |||||
1223  CTCAGGGCAACAGAGGCCCAACCAACCTCTCCAACAGATTAACGTGG 1272
297  LysPheProLeuThrArgProGlyTrpAspTrpAsnThrAlaGlyArg 313
   ||||| ||||| ||||| ||||| ||||| |||||
1273  GGTTCCTCTCGATGACCACTGGAGCTCAACACGCTCAAGAGTGG 1322
314  GluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAl 330
   ||| ||| ::||| ||||| ||||| ||||| |||||
1323  GAGAGTTGACAGTTTACTGCCGGCTTATGTGCAGGTCTCAAGGGAGC 1372
330  aSerArgArgProThrAsnLeuAlaLysValArgGluValMetGlnLysP 347
   ::||| ||||| ||||| ||||| ||||| ||||| |||||
1373  GGCAGAGATGCCCTTCATCTGGCAAGGTAAAGAGCTATGCAAGGGGC 1422
347  rOAsnGluProProSerValPheLeuGluArgLeuMetGluAlaPheArg 363
   || ||| ||||| ||||| ||||| ||||| ||||| |||||
1423  CAGCTGAACCCCGTGTCTTTCTAGAGATTAATGAGAGGCTTACGA 1472
364  ArgPheThrProPheAspProThrSerGlyAlaGlnLysAlaSerValAl 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1473  CGCTACACTCCCTTTGACCCGACTCGGAGGCTCAACACCCGGGTAGC 1522
380  aLeuAlaPheIleGlyLysSerAlaLeuAspIleArgLysLysLeuGlnA 397
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1523  TATGCTTTTATAGGCGATCATAGCAGCATTAAGCAAGCTTACAGA 1572
397  rGlyLeuGluGlyLeuGlnGlnAlaLeuArgAspLeuValArgGluAla 413
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1573  GACTGGAAGGTTATACATGCATGCTTACAGGACTT...AGAGAGCA 1618
414  GluLysValIleTyrTrpArgArgGluThrGluGluLysGlnGlnArg.L 430
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1619  GAGAAAGCTTTTACAGAGAGAAACAGAGAGCAAAAGAAAGAGAGAGA 1668
430  ysgLysGluArgGluGluArgGluGluArgArgAspArgArgGlnGlu 446
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1669  AGCATGAGAA...TAAAGAAAGAAATAGAGAGGATCAGAGGAGAG 1715
447  LysAsnLeuThrLysIleLeuAlaAlaValAlaGlnLysSerSerAr 463
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1716  AGAATTTAAGCAGAGATCTGCTGCAAGTAGTAGAAGAGCTGAGGAGAG 1765
463  gGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerG 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1766  AGCT.....AGATCTGAGAGTAGCAACAAAG 1791
480  LysAsnLeuGlyAsnArgThrPro..... 487
   || ||||| ||||| |||
1792  GGTACTCTGGGCAACAGGGAACCTTAGACTGTGAAGAAAACCCGTTGAG 1841
488  ....LeuAspLysAspGlnCysAlaTyrCysLysGlnLysGlyHisTr 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1842  AAAACTGTGAAGAGACCAATATGCTAGCTGAAAGAGAGGACATTTG 1891
1892  GCGCGCAGATGCGCCATGAGAGATGGGAATGCCCTAAGCTATATGATC 1941
519  euGluGluAsp 522
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1942  TCGAAGATGAT 1952
seq_name: gb_htc:AK004927
seq_documentation_block:
LOCUS AK004927 2935 bp mRNA htc 05-JUN-2001
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
ACCESSION AK004927
VERSION AK004927.1 GI:12836484
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
clone:lib.RIKEN full-length enriched mouse cDNA library
clone:1300007C21.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2935)
Carninci.P. and Hayashizaki.Y.
High-efficiency full-length cDNA cloning
Methods in enzymology. 303, 19-44 (1999)
99279253
MEDLINE 10349636
PUBMED 10349636
2 (bases 1 to 2935)
Carninci.P., Shibata.Y., Hayatsu.N., Sugahara.Y., Shibata.K.,
Itoh.M., Kono.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
genome research. 10 (10), 1617-1630 (2000)
20499374
JOURNAL
MEDLINE

```

PUBMED 11042159
 3 (bases 1 to 2935)
 SHIBATA, K., ITOH, M., AIZAWA, K., NAGAOKA, S., SASAKI, N., CARNINCI, P.,
 KONNO, H., AKIYAMA, J., NISHI, K., KITSUMI, T., TASHIRO, H., ITOH, M.,
 SUMI, N., ISHII, Y., NAKAMURA, S., HAZAMA, M., NISHINE, T., HARADA, A.,
 YAMAMOTO, R., MATSUMOTO, H., SAKAGUCHI, S., IKEGAMI, T., KASHIWA, K.,
 FUJIIAKE, S., INOUE, K., TOGAWA, Y., IZAWA, K., TANAKA, T., OHATA, E., WATANIKI, M.,
 YONEDA, Y., ISHIKAWA, T., OZAWA, K., TANAKA, T., MATSUURA, S., KAWAI, J.,
 OKAZAKI, Y., MURAMATSU, M., INOUE, Y., KIRA, A. and HAYASHIZAKI, Y.,
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome research. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4 (bases 1 to 2935)
 THE RIKEN Genome Exploration Research Group Phase II team and the
 FANTOM Consortium
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 2935)
 ARAKAWA, T., CARNINCI, P., FUKUDA, S., FUKUNISHI, Y., FURUKO, M.,
 HANAGAKI, T., HARA, A., HAYATSU, N., HIRAMOTO, K., HIROKA, T., HORI, F.,
 IMOTANI, K., ISHII, Y., ITOH, M., IZAWA, M., KATO, H., KAWAI, J.,
 KOJIMA, Y., KONNO, H., KONDA, M., KOYA, S., KURIHARA, C., MATSUYAMA, T.,
 MIYAZAKI, A., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OKAZAKI, Y.,
 OKIDO, T., OWA, C., SAITO, H., SAITO, R., SAKAI, C., SAKAI, K., SANO, H.,
 SASAKI, D., SHIBATA, K., SHIBATA, Y., SHINAGAWA, A., SHIRAKI, T.,
 SOGABE, Y., SUZUKI, H., TAGAMI, M., TAGAWA, A., TAKAHASHI, F.,
 TANAKA, T., TEJIMA, Y., TOYA, T., YAMAMURA, T., YASUNISHI, A.,
 YOSHIDA, K., YOSHINO, M., MURAMATSU, M. and HAYASHIZAKI, Y.,
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
 URL: http://genome.gscc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 Please visit our web site (http://genome.gscc.riken.go.jp/) for
 further details.
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5' GAGAGAGAGAGCGGCCCACTCGAGTTTCTTTTCTTTTVA 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. Second
 strand cDNA was prepared with the primer adapter of sequence [5'
 GAGAGAGAGAGATCCCAAGATCAATTAATTATTAACCCCCCCC 3']. cDNA was
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
 XhoI. Host: SOLR.
 Location/Qualifiers
 1. .2935
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="MGI:MGI:1897042"
 /db_xref="MGI:MGI:1914777"
 /clone="1300007C21"
 /sex="male"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 BASE COUNT 669 a 860 c 773 g 633 t
 ORIGIN
 alignment_scores:
 Quality: 1223.50 Length: 563
 Ratio: 3.245 Gaps: 17
 Percent Similarity: 66.963 Percent Identity: 48.135

alignment_block:
 US-09-171-553B-4 x AK004927 ..
 Align seg 1/1 to: AK004927 from: 1 to: 2935
 1 MetGlyInThrValThrProLeuSerLeuThrLeuAspHisTrpT 17
 |||||
 639 ATGGAGACAGACCGTAACACTACCCCCCGAGTCACTTAACCTTCACACACGAG 688
 17 hrgIuValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGly 33
 |||||
 689 GAGATGTCCAGCGCATTCATCCATCATGCTGTGTGATGTCCAGAGAGAG 738
 34 ProTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTr 50
 |||||
 739 CGCTGGGTACTTCTTCTTCGCCGAATGGCCCACTTCAATGTGGATG 788
 50 proSerGluGlyThrPheAsnSerGluLeileuAlaValLysAla 67
 |||||
 789 GCCTCAGATGATGTGCTTTATTTGGGTATATCTCTCAGCTCAAGTCTA 838
 67 leilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 83
 |||||
 839 GAGGTGTTTCTCTGGTCCCGACGAGACCCGGATCGAGTCCCATATATC 888
 84 LeuThrTrpGlnAspLeuAlaGluAspProPro 94
 |||||
 889 GTCACTGGAGAGGACCTCTCCAAACCCCTCTTACGACAGCTCCCG 938
 94 ..
 939 TCCGCCGCCGGGCTCTTGCGCAACCTCGGCCGATCGCCCTTAC 988
 95 ..
 989 CCGGCCCTTACCCCTCTATMAAGCC ..
 107 0 ..
 1030 CCAAGTTCTCCCTGATAGCGCGGACCTCTGACCTTCTCAGAGAG 1079
 117 ysaAnlyshsSerAlaGlnLysValGluProSerSerSerTrpLeuPro 133
 1080 ACCCGCCGCGTACGGA ..
 134 ArgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrcysSerProth 150
 1118 AGAGAGACGATGAMAAAGAGCGCGCCGCCAC ..
 150 rProLeuSerSerThrGlyCysGlyGluGlyThrSerAlaProProGly 167
 1151 ..
 167 LabProValValGluGlyProAlaAlaGlyThrArgSerArgGlyAla 183
 1170 CTCGCACTGCGGGGAGAGGAGGAGCCCT 1207
 184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrG 200
 1208 CCGGACGAGACCTCCACTCTCCAGGCAATTCCTCCGATG ..
 200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPhe 217
 1253 ..
 217 erSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
 1287 CCTCCTCTGACTT ..
 234 AspProGluArgLeuThrGlyLeuValGluSerLeuMetPheSerHisG 250
 1327 GACCCAGGTAAATTTGAGCGCTTGTGATGTGCTGCTCATCACCCAGCA 1376

250 nProthTTPAspaspCysgInglInleuengInthPheuphemrThmg 267
1377 GCCCACCTGGGAGCAGCTGTGAGAGTTGTTGGGAGCTTGGACCGAG 1426
267 IugIATrgGluIArgIleLeuLeuGluAlAlARgLySAsnValProGluYla 283
1427 AAGAAAGACAGCGGGTCTCTCTAGAGGCTAAGAAAGGCAATCCGGGCAAT 1476
284 AspGlyATgProthrgInleuGlnAsnGluIleAspMetGlyPheProLe 300
1477 GATGAGCGCCCGCACTCAGTTGGCCATAGAGCAATGCTGCTTTCCCT 1526
300 uThARgProGlyTTPAspTyrAsnThrIAGIugIArgGusSerLeuI 317
1527 TGAAGCGCCCGATGGGATTACACCACTACAGAGGTAGAGAACCACTAG 1576
317 ySIeTyrArgGlnAlAlLeuValAlAGIyLeuArGSLyAlAsARgArg 333
1577 TCCCTATGCGCAGTTCTCTTAGCGGGTCTCCAAACCGCGGAGAGAC 1626
334 ProthrsAsnLeuAlAsyValARgGluValAMetGlnGlyProAsnGluPr 350
1627 CCCACCAATTTGGCCAAAGTAAGGATPACCAGGAGCTTAATGAGTCTC 1676
350 oPserSerAlPheLeuGlnARgLeuMetGlnAlAPheArGATgPheThrP 367
1677 TCCTCAAGCCTTTTACAGAGACTCAAGAGGAGATATCCAGATCACTC 1726
367 ropheAspProthrsSerGluAlAGlnLySAlAsARyAlAlAlAlAhe 383
1727 CTTATGACCCCTGAGACCCAGGCGCAAGAAACAATGTGCTATGTCATTC 1776
384 IlegIyInserAlAlAsnAspIleArgLySIySLeuGInARgLeuGluI 400
1777 ATTCGGCAGTGTGCCCGGATATCGGGCGGAAGTTAGACGGTTAAAGA 1826
400 yLeuGlnGluAlAGIuLeuArgrAspLeuValARgGluAlAGIuLySValT 417
1827 TTTTAAAGAGCAAGACCTTAGAGACATTAGTAGAGGAGAGCTGAAGAATCT 1876
417 yTTPArTArGArGluThrGluGluGluLySGLuInARgLySGLuLySGLu 433
1877 TTTAATTAAGCGGAAMCCCGGAAGAAAGAGAGAAAGTATCAAGAGAA 1926
434 ArgGluGluARgGluGluArg..... 440
1927 ACAGAGCAAAAGAGAACCCCTAGGCGAGAGATGAGCAGAGAGACAA 1976
441 ...ArgspArgrArgGlnGluLySAsnLeuThrIySIleLeuAlAlay 456
1977 AGAAAGGAGACCGCAGAGACATAGAGATGAGCAAGCTCTGGCCACTG 2026
456 alValAGluGlyLySserSerArgrGluARgGluARgAspPheArGlySIle 472
2027 TAGTTAGTAGACAG..... 2040
473 ArgserGlyProArgrGInserGlyAsnLeuGlyAsnArgrThProLeuAs 489
2041 AGACAGGTT...AGACAGGGGGAGAG...CGAAGAGGCCCCCAACTTGA 2084
489 pLyAspGlnCysAlATyTCySlySGLuLySGLuYhISrPalArAsnC 506
2085 TTAAGACCAATATGCGCTACTGCAAAAGAAAGACACTGGGCTAAGACT 2134
506 ySProlLyS.....GlyAsnLyGlyProLyS 515
2135 GCCCAAAAGAACCCAGAGAGACCCCGAGAGACCGAG 2169
seq_name: gb_gss:BH063533

seq documentation block:
LOCUS BH063533 801 bp DNA GSS 18-JUL-2001

```

DEFINITION
  RPCI-24-370H8..TV RPCI-24 Mus musculus genomic clone RPCI-24-370H8.
  DNA sequence.
ACCESSION
  BH063533
VERSION
  BH063533.1
KEYWORDS
  GI:14878283
SOURCE
  house mouse.
  Mus musculus.
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus
  1 (bases 1 to 801)
AUTHORS
  Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akincet,B., Levins,M.,
  Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorjis,E.,
  Russell,D., de Jong,P. and Fraser,C.M.
  Mouse BAC End Sequences from Library RPCI-24
  Unpublished (1999)
  Other_GSSs: RPCI-24-370H8.TV
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-24. For BAC
  library availability, please contact Pieter de Jong
  (pdejong@email.cho.org). Clones may be purchased from BACPAC
  Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
  plate: http://www.tigr.org/tdb/Bac\_ends/mouse/Bac\_end\_intro.html
  Page: 370 row: H column: 8
  Seq primer: SP6
  Class: BAC ends.
FEATURES
  source
    location/Qualifiers
    1..801
    /organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="RPCI-24-370H8"
    /clone_lib="RPCI-24"
    /sex="Male"
    /cell_type="Spleen/Brain"
    /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
    RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
    library was cloned in the pTARBAC1 cloning vector at the
    BamH1 sites using MboI partially digested male C57BL/6J
    DNA."
BASE COUNT
  227 a 193 c 217 g 164 t
ORIGIN
  alignment_scores:
    Quality: 991.00 Length: 271
    Ratio: 4.164 Gaps: 1
    Percent Similarity: 87.823 Percent Identity: 71.587
  alignment_block:
    US-09-171-553B-4 x BH063533 ..
  Align seg 1/1 to: BH063533 from: 1 to: 801
  215 PropeserSerAlaAspLeuTyAAsnTrpLysThrAsnHisProProPh 231
    |||||.....||| |||.....|
  1 CCTTTTCCTCTCTGATTA..TATACTGGGAAACTAACCACTCCCTT 49
    |||||.....|
  231 eserGluAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheS 248
    |||||.....|
  50 CTCAGAGAACCCCTCTGGGCTTACTGGGCTTCCTGATCAGTCACTATGTCT 99
    |||||.....|
  248 eThiSgLnProThrTrpAspAspCysGlnGlnLeuLeuGlnTrhLeuPhe 264
    |||||.....|
  100 CCCATCAACCCACTGGGATGATGTACAGAGCTTTTGGCAGGTTCTTTTC 149
    |||||.....|
  265 ThrThrGluGluArgGluArgIleLeuLeuGlnAlaArgLysAsnValPr 281
    |||||.....|

```

150 ACCACGAGAAAGAGAAAGAAATCTGATGAGGCGAGAAAAATGTTCT 199
 281 OGILVALAAspGlyArgProThrInleuGlnAsnGluLeaspMetGlyP 298
 200 GGGAGAGAGCGGCACTGCCCTCCCTCCCTAACCCTGTCGAGCAGAGCTT 249
 298 heProLeuThrArgProGlyTrpAspTrpAsnThrAlaGluGlyArgGlu 314
 250 TCCCTTAAACCGCCCACTGGAGCTCAACACCGCGGAAAGGTAGGGGA 299
 315 SerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSe 331
 300 CCGCTTCCTGTTCTATCGCGGAGCTCTAGTGGAGGCTTCACAGAGCCAC 349
 331 rArgArgProThrAsnLeuAlaLysValArgGluValMetGlnGlyProA 348
 350 TAGACGGCCCAACCAATTGGCTAAGAGAGAGCTTTGAGGGGCGAGA 399
 348 sngLpProProSerValPheLeuGlnUArgLeuMetGluAlaPheArgArg 364
 400 CTGAACCAACCTCACTCTTCTGATCTGCTCATGTAGAGCATATAGAGAG 449
 365 PheThrProPheAspProThrSerGlnUAlaGlnLysAlaSerValAlaLe 381
 450 TACACCCCTTTGACCCCTCGTCAGAGGGGAGAAAGCCCGTACCAT 499
 381 uAlaIleIleGlyGlnSerAlaLeuAspLLeArgLysLysLeuGlnArgL 398
 500 GGCCTTCATGTCAGTCGCTCCGCTCCGCAATTAAAGAAAGCTGCAAGGC 549
 398 enGluGlyLeuGlnGlnUAlaGlnUArgAspLeuValArgGluAlaGlu 414
 550 TGGAGGGGCTCCACATCATACGCTCCCAAGATTTCATAAAGAACAGCAG 599
 415 LysValIleTyrTArgArgGluThrGluGluGluLysGluGlnArgLysG 431
 600 AAAGCTATCATATAGAGGAGAACCAAGAGAGAGAGAGAGAGAGAGAA 649
 431 uLysGluArgGluGluGluGluGluGluGluGluGluGluGluGluGlu 448
 650 GAAAGAAATGAGAGAGAGAGAAATAGAGGAGATGGGCTCGAGAGAGAA 699
 448 sLeuThrLysIleLeuAlaValAlaValGluGlyLysSerSerArgL 464
 700 ATTGAGTAAATTTTGGCCGAGTTGTAATGATAGA..... 737
 465 ArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAs 481
 738:CACTCAGGAAAGGTAAACAGAGACT 763
 481 nLeuGlyAsnArg 485
 764 TCTGGGCAACAGG 776

seq_name: gb_gss: A2318021

seq_documentation_block:

LOCUS A2318021 729 bp DNA GSS 29-SEP-2000

DEFINITION IM0036004R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION A2318021

VERSION A2318021.1 GI:10367383

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 729)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL Plasmid inserts
 Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunne@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0036 row: 0 column: 04
 Seq primer: CACACAGCAACACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 729.

FEATURES

source

1..729

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M036004"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/shares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (q11473211419b1AF129072.1), a copy number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 196 a 193 c 182 g 158 t

ORIGIN

alignment_scores:

Quality: 946.00 Length: 239

Ratio: 4.359 Gaps: 2

Percent Similarity: 90.795 Percent Identity: 75.314

alignment_block:

US-09-171-553B-4 x A2318021 ..

Align seg 1/1 to: A2318021 from: 1 to: 729

194 LeuProLeuArgThrTyr...GlyProProMetProGly...GlyGlnLeu 208
 11 CTGCGCTTAAAGCACAATGGAGGAGGCGACGCGCAGACCTAATGATCT 60
 208 uGlnProLeuGlnTyrTrpProPheSerAlaAspLeuTyrAsnTrpL 225
 61 CATTCCTTTACATACAGTGGCTTTCTCTTCGATTATATATATATGA 110
 225 ysthrAsnHisProProPheSerGluAspProGlnArgLeuThrGlyLeu 241
 111 AAACCTAACACCCCTCCCTTCAGAGAACCCCTTGCGCTTACAGGGCTC 160
 242 ValGlnSerLeuMetPheSerHisGlnProThrTrpAspAspGlyGln 258
 161 CTGAGTCATTAATGTTCTCCATCAACCACTTGGATGATTTTCAGCA 210

258 nteuLeuGlInThrLeuPheThrThcGlInLuarGluArgIleLeuG 275
 211 GCTTTGACAGGTTCTTTCACACAGAAAGAGAAAGAAATCCGATGG 260
 275 luAlaArgLysAsnValProGlyAlaAspGlyArgProThrcGlnLeuGln 291
 261 AGCGGAGAAAAATGTTTACAGAGAGAGACACACACCCCTCCCT 310
 292 AsnGluIleAspMetGlyPheProLeuThrArgProGlyTyrAspTyrAs 308
 311 AACCTGTCGAGAGGCTTCCCTTGAAACGCCCAACTGGGACTACAA 360
 308 nThrAlaGluArgGluSerLeuLysIleTyrArgGlnAlaLeuValA 325
 361 CACCGCGAGAGGTAGGGAGCGCTCTGTCTATCGCGGACTCTAGTGG 410
 325 laGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArg 341
 411 CAGGCTCAGAGAGAGCGCTAGACGGCCCAATTTGGCTTAGGTAAAG 460
 342 GluValMetGlnGlyProAsnGluProSerValPheLeuGluArgLe 358
 461 GAGGCTCTGACAGGGGACAGCTGACACACCCCTCAGTCTTCTGAGCGTCT 510
 358 uMetGluAlaPheArgArgPheThrProPheAspProThrSerGluAlaG 375
 511 CATGGAGGCTATTAAGAGGTACACCCCTTTGACCCCTGCTCAGAGGGGC 560
 375 lAlaLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIle 391
 561 AAAAAAGCCCTTACGCTTACGCTTCAATTTGTCAGTCTCGCTCCGACATTT 610
 392 ArgLysLysLeuGlnArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 408
 611 AAGAAAGAGCTGCAAGGCTGGAGGGGCTCCAGATCATACGCTCCAGAA 660
 408 pleuValArgGluAlaGluLysValTyrTyrArgArgGluThrcGlnG 425
 661 TTTTACTTAAAGACAGAGAAAGTCTATCATPAGAGGAGAAACAGAAAG 710
 425 lLysGluGlnArgLys 430
 711 AGAGCGAGAGAGAGAG 727

seq_name: gb_gss:BH121259

seq_documentation_block: 679 bp DNA 19-JUL-2001
 LOCUS BH121259
 DEFINITION RPCI-24-374G11.TV RPCI-24 Mus musculus genomic clone RPCI-24-374G11
 , DNA sequence.
 ACCESSION BH121259
 VERSION BH121259.1 GI:14964771
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 679)
 Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akhret,B., Levins,M.,
 Tesgaye,G., Geer,K., Malek,M., Shvartsbeyn,A., Gebregorjais,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other_GSSs: RPCI-24-374G11.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/Bac_ends/mouse/bac_end_intro.html
 Plate: 374 row: G column: 11
 Seq primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1. 679
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-374G11"
 /clone_11b="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTRABAC1; Site1: BamHI; Site2: BamHI;
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTRABAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 194 a 167 c 174 g 144 t
 ORIGIN

alignment_scores:
 Quality: 916.00 Length: 226
 Ratio: 4.425 Gaps: 1
 Percent Similarity: 91.593 Percent Identity: 76.991

alignment_block:
 US-09-171-553B-4 x BH121259 ..

Align seg 1/1 to: BH121259 from: 1 to: 679

214 TrrProPheSerSerAlaAspLeuTyrAsnTrrLysThrAsnHisProp 230
 1 TGGCTTTTTCCTCTGATTATATACAAATGGGAAACATACACCCCTTC 50
 230 rOPheSerGluAspProGlnArgLeuThrcGlyLeuValGluSerLeuMet 246
 51 CCTTCTCAGAGAAACCCCTCTGAGACTTACTGGCTCTCTGAGTCACTTAG 100
 247 PheSerHisGlnProThrTrrPaspAspCysGlnGlnLeuGlnThrLe 263
 101 TTCTCCATCACCACCACTTGAGATGATTGTCACACACTTTTGACAGTTCT 150
 263 upheThrThcGlnGluArgGluArgIleLeuLeuGlnAlaArgLysAsn 280
 151 TTTTACCACAGAGAAAGAGAAAGAAATCCTGATGAGGCGGCAAAAAATG 200
 280 alProGlyAlaAspGlyArgProThrcGlnGlnGlnGlnGlnGlnGlnGln 296
 201 TTCTGGAGAGAGAGAGGACACCCACCTGCTCCCTCAACTCGTGGACGAG 250
 297 GlyPheProLeuThrArgProGlyTrrPaspTyrAsnThrAlaGluLys 313
 251 GCTTTTCCCTTGAACACACCCCACTGGAGCTACACACCGGAGAGCTAG 300
 313 gGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyA 330
 301 GGGAGCGCTCTCTGTATCGCGGACTCTAGAGGAGGCTTCAGAGGAG 350
 330 laSerArgArgProThrAsnLeuAlaLysValArgGluValMetGlnGly 346
 351 CCGCTAGACGGCCCACTTTGGCTTAGGTAAAGAGGCTTTGACAGGG 400
 347 ProAsnGluProProSerValPheLeuGluArgLeuMetGluAlaPheAr 363
 401 CAGACTGAAACACCATCAGCTCTCTTGAGCGCTATAGGAGCATATAG 450
 363 gatGpHeThrProPheAspProThrSerGluAlaGlnLysAlaSerValA 380
 451 GAGATACACCCCTTTTACCCCTTTCAGAGAGGCGAGAGACCGCTGATG 500

seq_documentation_block: 644 bp DNA GSS 19-JUL-2001
 LOCUS BH124013
 DEFINITION RPI-24-230P16.TV RPI-24 Mus musculus genomic clone RPI-24-230P16
 , DNA sequence.
 ACCESSION BH124013
 VERSION BH124013.1 GI:14967525
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 644)
 Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levin, M.,
 Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPI-24
 Other GSSs: RPI-24-230P16.TV
 CONTACT: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
 Page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 230 row: P column: 16
 Seq primer: T7
 Class: BAC ends.
 FEATURES
 source Location/Qualifiers
 1..644
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-24-230P16"
 /clone_id="RPI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pPARBAC1; Site:1: BamHI; Site:2: BamHI;
 RPI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pPARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."
 BASE COUNT 176 a 162 c 169 g 137 t
 ORIGIN
 alignment_scores: Quality: 846.00 Length: 213
 Ratio: 4.383 Gaps: 0
 Percent Similarity: 90.610 Percent Identity: 75.117
 alignment_block:
 US-09-171-553B-4 x BH124013 ..
 Align seg 1/1 to: BH124013 from: 1 to: 644
 223 AaTTPPLVsthrAsnHisProPheserGluaspproGluArgLeuTh 239
 |||||||
 3 AATTGGAACAACTAACCCCTCCCTTCACAGAACCCCTCGACTTAC 52
 |||||||
 239 rGlyLeuValGluSerLeuMetPheSerHisGlnProThrTrpAspSC 256
 |||||||
 53 TGGGGCTCCCTGAGTCATTGTTCTCCATCAACCCACTGGGATGATT 102
 |||||||
 256 ySGlnGlnLeuLeuGlnThrLeuPheThrThrGlnGlnArgGlnArgLe 272
 |||||||
 103 GTGACAGAGCTTTGACAGTCTTTTACACAGAGAAAGAGAAAGATC 152
 |||||||

273 LeuLeuGlnAlaArgLysAsnValProGlyAlaAspGlyArgProThrG 289
 |||||||
 153 CTGATGAGAGCGGATTTTTTTGTTCTGTGGAGAGAGCGGACACCACTGC 202
 |||||||
 289 nLeuGlnAsnGlnLeuAspMetGlyPheProLeuThrArgProGlyTrpA 306
 |||||||
 203 CCTCCCTAACCTCGTTGACGAGGCTTTTCCTTAACCCGCCCAACTGG 252
 |||||||
 306 sPTyrAsnThrAlaGlnValArgGluSerLeuLysIleTyrArgGlnAla 322
 |||||||
 253 ACTACACACCGCGGAGAGAGGAGGAGCGCTCTTGTCTATACCGGACT 302
 |||||||
 303 CTAGAGGAGAGCTCTCAGAGAGCGCGCTAGAGCGCCACCAATTGGCTAA 352
 |||||||
 339 sValArgGluValMetGlnGlyProAsnGluProProSerValPheLeuG 356
 |||||||
 353 GGTAAAGAGAGCTCTTGACGAGGCGAGACTGAACCACTCTTCCTTG 402
 |||||||
 356 lUArgLeuMetGluAlaPheArgArgPheThrProPheAspProThrSer 372
 |||||||
 403 AGCGTCTATGAGGAGCATATGAGATACACCTTTTGACCCCTTGCTCA 452
 |||||||
 373 GluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyInSerAlaLe 389
 |||||||
 453 GAGGGGAGAGAGCGCGCTAGCATGGCCCTCTATTGGTCAATCGCTCC 502
 |||||||
 389 uAspIleArgLysLysLeuGlnArgLeuGlnGlyLeuGlnGlnAlaGlu 406
 |||||||
 503 CGACATTAAAGAAAGAAAGCTGACGAGGCTGAGGGCTCCAGATCATACG 552
 |||||||
 406 eUArgAspLeuValArgGluAlaGlnLysValTyrTyrArgArgGluThr 422
 |||||||
 553 TTCCAAGATTGTTAAAGAACAGACAGAAAGCTCTATCAAAAGGAACA 602
 |||||||
 423 GluGlnGlnLysGlnGlnArgLysGlnLysGlnArgGlu 435
 :::::::::::::::::::::
 603 CGAGAGAGAGAGCGAGAGAGAGAGAGAGAAAGAAATGAG 641
 :::::::::::::::::::::
 seq_name: gb_gss:AZ420279
 seq_documentation_block: 617 bp DNA GSS 03-OCT-2000
 LOCUS AZ420279
 DEFINITION IM0198K05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0198K05 F, DNA sequence.
 ACCESSION AZ420279
 VERSION AZ420279.1 GI:10544292
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 617)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tinley, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 CONTACT: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 1000 Std Error: 0.00
 Plate: 0198 row: K column: 05
 Seq primer: CGTTGTAACGACGCGCAGT

Class: plasmid ends
High quality sequence stop: 617.
Location/Qualifiers
1. 617

FEATURES
SOURCE
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0198K05"
/clone_id="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (9114732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 138 a 158 c 155 g 166 t
ORIGIN

alignment_scores:
Quality: 845.50 Length: 212
Ratio: 4.450 Gaps: 1
Percent Similarity: 89.623 Percent Identity: 76.887

alignment_block:

US-09-171-553b-4 x AZ420279/rev ..

Align seg 1/1 to reverse of: AZ420279 from: 1 to: 617

```

219 AlAspLeuTyrAsnTrpLysThrAsnHisProPheSerGluAsp 235
      ::::::::::::::::::::::::::::::::::::::::::::
615 TCGGATCTTACATGGAAACCTAATCACACCTTCTCTGAAACCC 566
235 oGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGlnPro 252
      ::::::::::::::::::::::::::::::::::::::::::::
565 CTCAGGGTTACTGACTGCTCTAGAGTCCCGATGTTTCCACACCCCA 516
252 hTrTAspAspCysGlnGlnLeuGlnThrLeuPheThrThrGluGln 268
      ::::::::::::::::::::::::::::::::::::::::::::
515 CTGGGATGACTGCATCAACAGCTTTTACAGGTCTCTTCAACAGGAAGA 466
269 ArgGlnArgGlyLeuLeuGlnValArgGlyAsnValProGlyAlaAsp 285
      ::::::::::::::::::::::::::::::::::::::::::::
465 CGGGAAGAAATCTTCTGGAGGCTAGAAAAAATGTCTGGGAGACATGG 416
285 yArGProThrGlnLeuGlnAsnGlnIleAspMetGlyPheProLeuThra 302
      ::::::::::::::::::::::::::::::::::::::::::::
415 GAGCCCAACACACCTTCTTAACCTTAATGTGATGAAGCTTTTCCCTTGA 366
302 rGpProGlyTyrAspTyrAsnThrAlaGlnGlyArgGlnSerLeuLysIle 318
      ::::::::::::::::::::::::::::::::::::::::::::
365 GCCCAAACTGGGACTACATACAGCAAGAGTGGAGGCTCTCCGCTGCTC 316
319 TyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgProThr 335
      ::::::::::::::::::::::::::::::::::::::::::::
315 TATGCCCGGACTAGTGGCGGCTCTTAAAGGGGCTGACAGAGGCCAC 266
335 rAsnLeuAlaLysValArgGlnValMetGlnGlyProAsnGlnProPro 352

```

```

265 CAATCTGGCAAGGTAAGAGAGTCTTGCAGACGACGAGCCACCT 216
352 eValPheLeuGlnArgLeuMetGlnAlaPheArgArgPheThrProPhe 368
      ::::::::::::::::::::::::::::::::::::::::::::
215 CGGCTCTTCTAGAACGGCTAATGAGGCTATAGAGGTATACCCGCTTC 166
369 AspProThrSerGlnAlaGlnLysAlaSerValAlaLeuAlaPheIleG 385
      ::::::::::::::::::::::::::::::::::::::::::::
165 GATCTTCATCGAGGGGACAGCAACGCTTCTCTGCTTATCATTTGG 116
385 yGlnSerAlaLeuAspIleArgGlyLysLeuGlnArgLeuGlnGlyLeuG 402
      ::::::::::::::::::::::::::::::::::::::::::::
115 CCAGTCACCTCCCGATATATAAAGAGCTCAAGACCTCGAGGGCTCC 66
402 InGlnAlaGlnLeuArgAspLeuValArgGlnAlaGlnLysValTyr 418
      ::::::::::::::::::::::::::::::::::::::::::::
65 AAGATTATACCTTTCGAGGATCTAGTGAAGAGCAAGAAAGTGTCTTA 16
419 ArgArgGlnThrGlnGlnGlnGlnGlnArgLys 430
      ::::::::::::::::::::::::::::::::::::::::::::
15 AGA.....GAGAGACGAGAAG 1

```

seq_name: gb_gss:AZ986580

seq_documentation_block: 677 bp DNA GSS 27-APR-2001

LOCUS AZ986580 677 bp DNA GSS 27-APR-2001
DEFINITION 2M0268P18R Mouse 10kb plasmid U06C2M library Mus musculus genomic
clone U06C2M0268P18 R, DNA sequence.

ACCESSION AZ986580
VERSION AZ986580.1 GI:13857807

KEYWORDS GSS.
SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 677)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Published (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std error: 0.00
Plate: 0268 row: P column: 18
Seq primer: CACACAGCAACAGCTATGACCC
Class: plasmid ends
High quality sequence stop: 677.

FEATURES
SOURCE

1. 677
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C2M0268P18"
/clone_id="Mouse 10kb plasmid U06C2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4

alignment_scores:

Quality: 827.00 Length: 223
 Ratio: 4.241 Gaps: 3
 Percent Similarity: 87.444 Percent Identity: 73.991

alignment_block:

us-09-171-553b-4 x AZ862103/rev ..

Align seg 1/1 to reverse of: AZ862103 from: 1 to: 677

```

189 AspgluileuAlaIleuProleuArgThr...GlyProProMetPr 204
|||||...|||||...|||||...|||||...|||||...
660 GATTCAACAGTTCCTGACCTTAGACACATGTGGAGGCGCAGCGCC 611
204 GGLY...GlyGlnLeuGlnProleuGlnTyrTyrProheseSerAla 220
|||||...|||||...|||||...|||||...|||||...
610 AGGACCTAATGATTCATTCCTTACATCTGCGCTTTCTCTCTG 561
220 SPLLeuTyrAsnTyrPlyThrAsnHisPropheSerGluAspProGln 236
|||||...|||||...|||||...|||||...|||||...
560 ATTATATATTTGAAACCTAACCCCTCCCTCTCAGAGAACCCCTCT 511
237 ArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGlnProThr 253
|||||...|||||...|||||...|||||...|||||...
510 GGGCTTACTGGGCTTCCTGAGTCACTATGTTCTCCCATCAACCCACTG 461
253 PASpAspGlyGlnGlnLeuGlnThrLeuPheThrThrGluGluArg 270
|||||...|||||...|||||...|||||...|||||...
460 GGATATATGTCAGCAGCTTTGACAGTCTTTTACACACAGAACAGAG 411
270 LuArgIleLeuLeuGluAlaArgLysAsnValProGlyAlaAspGlyArg 286
|||||...|||||...|||||...|||||...|||||...
410 AAAGATCTCTGATGAGGCTAGAAAAAATTTCTAGAGAGAGAGGACACA 361
287 ProThrGlnLeuGlnAsnGlnIleAspMetGlyPheProLeuThrArgPr 303
|||||...|||||...|||||...|||||...|||||...
360 CCCACTGCTCTCCCTAACCTCGTAGAGGCTTTCCCTTGAAACCGGCC 311
303 GGLYTPAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyr 320
|||||...|||||...|||||...|||||...|||||...
310 CAACTGGAGCTTACACACCGCGGAGGTAGGGGACACCTCTCTTATC 261
320 rGlnAlaIleLeuValAlaGlyLeuArgLysAlaSerArgArgProThrAsn 336
|||||...|||||...|||||...|||||...|||||...
260 GCCGACTCTAGTGGCAGAGTCTCAGAGGGC..... 230
337 LeuAlaLysValArgGluValMetGlnGlyProAsnGluProProSerVa 353
|||||...|||||...|||||...|||||...|||||...
229 CTGGCTAGAGTAAGAGAGGCTCTCAGAGGCGACACTGACACCCCTCAGT 180
353 lPheLeuGluArgLeuMetGluAlaPheArgArgPheThrProheseSP 370
|||||...|||||...|||||...|||||...|||||...
179 CTTCCTTGAGCGTCTCATGAGGATATAGAGGCTACACCCCTTTTAC 130
370 rGlnrSerGluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGln 386
|||||...|||||...|||||...|||||...|||||...
129 CCTGCTCAGAGGCGCAGAAAGCTCTGATGACATGCGCTTCATTGGTCAG 80
387 SerAlaLeuAspIleArgGlyLysLeuGlnArgLeuGlnGlyLeuGlnG 403
|||||...|||||...|||||...|||||...|||||...
79 TCCCGCTCCGACATTAGAAAAAGCTCAAGAGGCTGAGAGGCTCCACAGA 30
403 uAlaGluLeuArgAspLeu 409
|||||...|||||...
29 TCATACGCTCCAGATTTA 11
seq_name: gb_est2:BG089942

```

seq_documentation_block: 645 bp mRNA EST 26-JAN-2001
 LOCUS BG089942
 DEFINITION mab8d06.x1 NCI_CGAP.SP2 Mus musculus cDNA IMAGE:3977506 3'

similar to TR:Q9WLL1 Q9WLL1 GAG PROTEIN. [1] ;, mRNA sequence.

ACCESSION BG089942
 VERSION BG089942.1 GI:12572505
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 TITLE 1 (bases 1 to 645)
 JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Other ESTs: mab8d06.y1
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail@nih.gov
 Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

MG1:1477538
 Seq primer: -40UP from Gibco
 High quality sequence stop: 422.
 Location/Qualifiers

FEATURES

source

1..645
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:3977506"
 /clone_lib="NCI_CGAP.SP2"
 /tissue_type="NK cells (flow-sorted)"
 /lab_host="DH10B (TI-resistant)"
 /note="Organ: spleen; Vector: pCMV-SPORT6 (Life Technologies); mRNA made from flow-sorted NK cells, CDNA made by oligo-dT priming. Directionally cloned. Average insert size 1.5 kb. Primary library, non-amplified. CDNA library preparation: David B. Kitzman, Ph.D."
 BASE COUNT 141 a 177 c 150 g 176 t 1 others
 ORIGIN

alignment_scores:

Quality: 810.50 Length: 215
 Ratio: 4.311 Gaps: 1
 Percent Similarity: 87.442 Percent Identity: 75.349

alignment_block:

us-09-171-553b-4 x BG089942/rev ..

Align seg 1/1 to reverse of: BG089942 from: 1 to: 645

```

216 PheSerSerAlaAspLeuTyrAsnTyrPlyThrAsnHisPropheSe 232
|||||...|||||...|||||...|||||...|||||...
642 TTTTCTCTCTGCTACCTTTACAAATGAAGGTC...ATCCCCCTTTT 596
232 rGluAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSer 249
|||||...|||||...|||||...|||||...|||||...
595 CAGAAACTGCAAGAGGACTGAGTGGTTGATGATTAATGATTTTAC 546
249 lAslProThrTyrPAspAspGlyGlnGlnLeuGlnThrLeuPheThr 265
|||||...|||||...|||||...|||||...|||||...
545 ACCAGCCGAGCTTGGATGACCTGCCAGCAGCTTCTGAGACTCTATTTCACA 496
266 ThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProG 282
|||||...|||||...|||||...|||||...|||||...
495 ACCGAGGAGAGAGAGATTTCTCTCGAGGCTCGAAAAATGTCGAGACA 446
282 yAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheP 299
|||||...|||||...|||||...|||||...|||||...
445 CGAGGCTGGGCGCCCTGTCAAACCTCAGCTGATGATGATGAAGATTTC 396

```


page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html

Plate: 238 row: P column: 4
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers

1..820
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPECI-23-238P4"
/clone_lib="RPECI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 236 a 186 c 247 g 151 t
ORIGIN

alignment_scores:
Quality: 797.50 Length: 253
Ratio: 3.853 Gaps: 5
Percent Similarity: 81.818 Percent Identity: 65.217

alignment_block:

US-09-171-553B-4 x AZ708792 ..

Align seg 1/1 to: AZ708792 from: 1 to: 820

```

280 ValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluLeuAspMe 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1  GCTTGGGGAAGAGGAGCTCCACCACTCTTAACCTTATTGATGA 50
296 tGlyPheProLeuThrArgProGlyTyrAspTyrAsnThrAlaGluGlyA 313
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
51 AGCTTTCTCTTGACCCGCCCAACTGGACTACATACAGCAGAAAGTA 100
313 tGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGly 329
    ||| ||| : : : : : : : : : : : : : : : : : : : : :
101 GGGAGCGCTCTCGTCTATCGCCGACTAGTGGCGGCTCAAAAGCG 150
330 AlaSerArgArgProThrAsnLeuAlaLysValAlaArgGluValMetGln 346
    ||| : : : : : : : : : : : : : : : : : : : : :
151 GCTGCAGACGCGCCCTTCTTCTTGCAAGGTAGAGAGGCTTGCAGGG 200
346 yProAsnGluProProSerValPheLeuGluArgLeuMetGluAlaPheA 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 ACCAGCGAGCCACCTCGTCTTAGAAGCGCTAATGAGGCGCTATA 250
363 tGArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerVal 379
    ||| : : : : : : : : : : : : : : : : : : : : :
251 GGAGGTATACCCCGTTCATCTTCATCGGAGGGGCAAGCAAGCAAGCTGT 300
380 AlaLeuAlaPheIleGlyLysSerAlaLeuAspIleArgLysLysLeuG 396
    ||| : : : : : : : : : : : : : : : : : : : : :
301 GCTATGGCTTTTCATTCAGAGAGAGAGAGAGATGAAAAAGAAAGTCA 350
396 nArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluA 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 AAGACTCGAGGGGCTCCAGATTTACTTTCAGAGATCTAGTGAAGGAAG 400
413 laGluLysValTyrTyrArgArgGluThrGluGluGluLysGluGlnArg 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 CAGAGAAAGTGTTCATTAAGAGAGAGAGAGATGAAAAAGAAAGAAAGA 450
430 LysGluLysGluArgGluGluGluGluGluGluGluGluGluGluGlu 446
    : : : : : : : : : : : : : : : : : : : : :
451 GAGAAAAAGAGGCAAGAGACAGGAGATGACGGGACCGTGCACAGGA 500

```

```

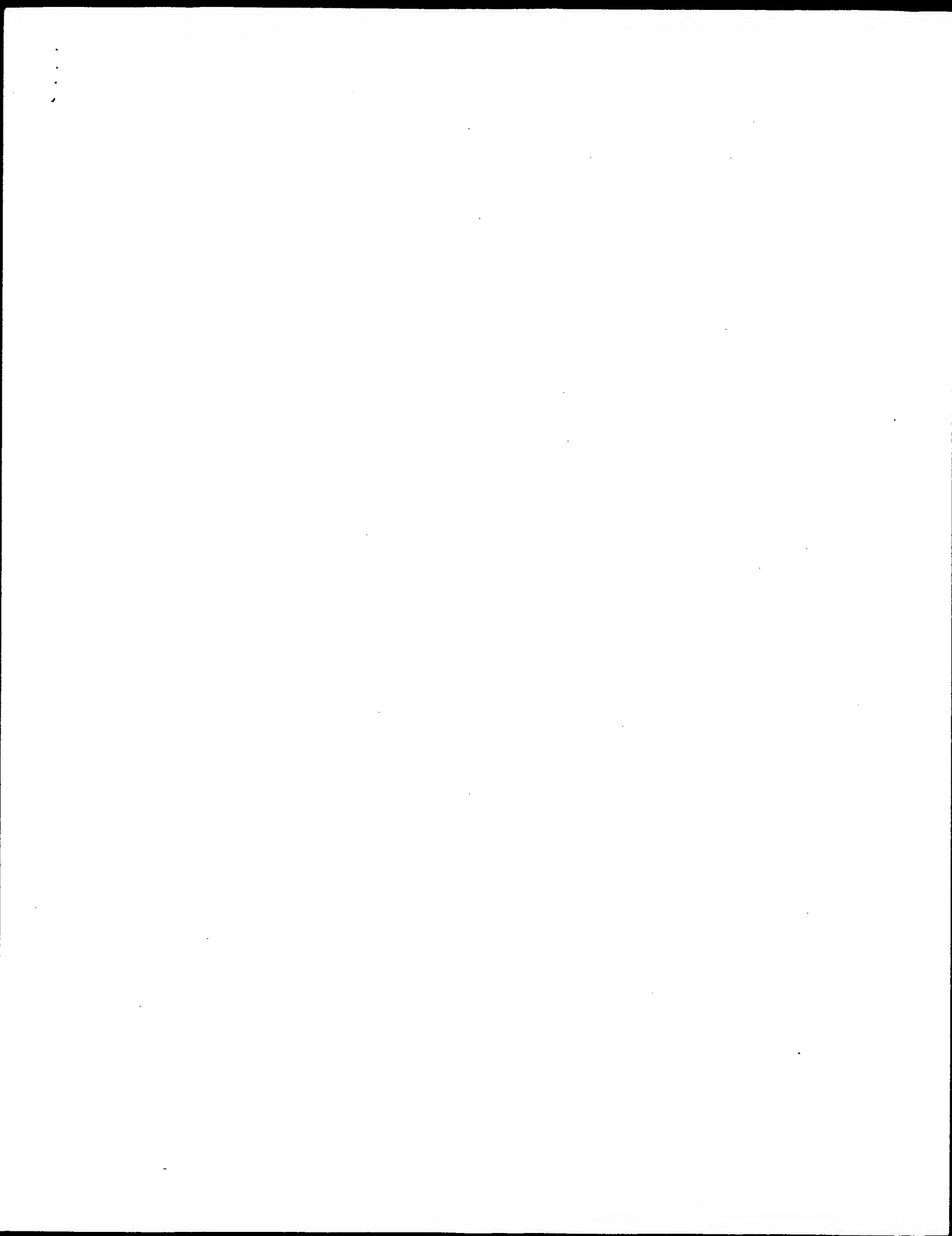
446 uLysAsnLeuThrLysIleLeuAlaAlaVal...GluGlyLysSerS 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
501 AAGAAACCTGACAGAGATTTCTGGCCGCACTGTAAGAGAGACAGTAA 550
462 eArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGln 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
551 GGAGAGATAGG..... 561
479 SerGlyAsnLeuGlyAsn.....ArgThrPr 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
562 TCAGGGAACCTGGCACTGGCGAGAGACGAGATTGCAAGACCCAAACC 611
487 o...LeuAspLysAspGlnCysAlaTyrCysLysGluLysGlyHisTrpAl 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
612 CAGTTGAAAGAGACAGTGTGGTATTGCAAAAGAAAGGAGCATTTGGGT 661
503 aArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuG 520
    : : : : : : : : : : : : : : : : : : : : :
662 TCGGGAATGCCCAAGAGAGA.....GCCAAGGTGCTGGCCCTCG 701
520 LngLusP 522
    : : : : : |||
702 ATCAAGAT 709

```

Mon Feb 25 07:44:20 2002

us-09-171-553b-4.feb22std.rst

Page 17



OM of: US-09-171-553b-5 to: EST:* out_format: pfs
 Date: Feb 24, 2002 8:31 AM
 About: Results were produced by the genecore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frimex.p2n.model
 -O=/cgm2.1/USPTO.spool/US09171553/runat_22022002_122825_11337/app-query.fasta.1.3312
 -DB=EST -OEM=fastap -SUFFIX=feb22std.rst -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.500
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FCGAPOP=6.000 -FCGAPEXT=7.000 -START=1 -MATRIX=biosum62
 -DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
 -TRANS=human4.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pls
 -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USPR=US09171553 @CGN1.1 9008 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-171-553b-5
 Database: EST.*
 Database sequences: 11351937
 Database length: 1077921985
 Search time (sec): 3515.700000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_gss:AZ636636	- 1039.00	1367.71	3.7e-78	685	AZ636636 IM0495002R Mouse 10kb
gb_gss:BH090616	+ 1011.50	1524.37	9.6e-76	791	BH090616 RPT-24-35614.TV RPT
gb_est2:B1107577	+ 1004.00	1512.85	4.2e-75	799	B1107577 602894023F1 NCI_CGAP_L
gb_gss:BH056750	+ 999.50	1505.20	1.1e-74	864	BH056750 RPT-24-94C16.TVB RPT
gb_gss:AZ980516	- 993.00	1505.35	1.1e-74	882	AZ980516 2M0257006R Mouse 10kb
gb_est2:B1152889	+ 993.00	1495.95	3.7e-74	813	B1152889 602918048F1 NCI_CGAP_L
gb_gss:AZ625666	+ 988.00	1490.01	7.9e-74	690	AZ625666 IM0469120R Mouse 10kb
gb_gss:AZ717731	+ 981.00	1477.62	3.9e-73	817	AZ717731 RPT-24-15511.TV RPT
gb_gss:AZ715956	+ 972.50	1465.04	1.9e-72	789	AZ715956 RPT-24-155A1.TV RPT
gb_est2:BF784826	+ 967.00	1454.49	7.5e-72	945	BF784826 60211174F1 NCI_CGAP_L
gb_gss:BH106644	+ 966.50	1455.24	6.8e-72	842	BH106644 RPT-24-352A17.TVB RPT
gb_gss:AZ715722	+ 961.00	1447.99	1.7e-71	755	AZ715722 RPT-24-154N5.TV RPT
gb_gss:AZ827524	+ 960.50	1448.01	1.7e-71	700	AZ827524 2M0104E02F Mouse 10kb
gb_gss:AZ982997	- 952.00	1435.56	8.5e-71	668	AZ982997 2M0264E02F Mouse 10kb
gb_hic:AK004796	+ 950.00	1416.27	1.0e-69	3234	AK004796 Mus musculus adult ma
gb_hic:AK004571	+ 947.00	1411.52	7.1e-70	3293	AK004571 Mus musculus adult ma
gb_gss:BH083773	+ 941.00	1419.05	7.1e-70	653	BH083773 RPT-24-293G13.TV RPT
gb_gss:BH083773	+ 939.50	1415.48	1.1e-69	740	AZ631391 IM0485106R Mouse 10kb
gb_gss:BH122448	+ 938.50	1413.44	1.4e-69	778	BH122448 RPT-24-330J18.TV RPT
gb_gss:AZ353894	+ 931.00	1404.04	4.8e-69	640	AZ353894 IM0099202R Mouse 10kb
gb_est1:AL1134888	+ 929.00	1401.12	7.0e-69	632	AL1134888 DRFP/6240910.T1 762
gb_gss:AZ238859	+ 922.50	1390.47	2.8e-68	632	AZ238859 RPT-23-87K22.TV RPT
gb_gss:AZ866393	- 918.50	1389.73	6.5e-68	725	AZ866393 2M0176E18R Mouse 10kb
gb_gss:AZ866393	- 917.00	1381.79	8.4e-68	701	AZ866393 2M0165H23R Mouse 10kb
gb_gss:AZ860071	+ 917.00	1372.64	2.7e-67	702	AZ860071 2M0249C06F Mouse 10kb
gb_gss:AZ860071	+ 911.00	1370.88	3.4e-67	719	BH042410 RPT-24-301E24.TV RPT
gb_gss:BH042410	+ 908.00	1368.03	4.9e-67	705	AZ337037 IM0067L03R Mouse 10kb
gb_gss:AZ237037	+ 902.50	1359.47	1.5e-66	718	AZ618408 IM0450C07F Mouse 10kb
gb_gss:AZ237037	+ 902.00	1358.25	1.7e-66	751	BH092018 RPT-24-35616.TV RPT
gb_gss:BH092018	+ 900.50	1356.87	2.1e-66	688	AZ825369 2M0100I04R Mouse 10kb
gb_gss:AZ825369	- 898.00	1357.25	2.0e-66	616	AZ962118 2M0230N14R Mouse 10kb
gb_gss:AZ825369	- 898.00	1357.25	2.0e-66	780	X83313 MMR516R5R Mouse cDNA Mus
gb_gss:AZ825369	- 891.00	1343.17	1.2e-65	639	AZ644009 IM0505011F Mouse 10kb
gb_gss:AZ825369	+ 891.00	1342.37	1.3e-65	691	BH034484 RPT-24-311D3.TV RPT
gb_gss:BH034484	+ 891.00	1340.17	1.8e-65	861	BH1143830 60290701F1 NCI_CGAP_L
gb_gss:BH041228	+ 890.00	1341.54	1.5e-65	946	BH041228 RPT-24-329L23.TVB RPT
gb_est2:BF162779	+ 889.50	1335.65	3.1e-65	987	BF162779 602918918F1 NCI_CGAP_L
gb_est2:BF162779	+ 888.50	1329.57	6.8e-65	717	BF162779 601769243F1 NCI_CGAP_L
gb_gss:BH074726	+ 883.00	1328.58	7.8e-65	734	BH074726 RPT-24-331M11.TVB RPT
gb_gss:BH028043	+ 883.00	1328.53	7.8e-65	812	BH028043 RPT-24-247H6.TV RPT

gb_gss:AZ687994 - 881.00 1327.18 9.2e-65 689 AZ687994 RPT-24-173A19.TV
 gb_gss:AZ403060 - 880.00 1325.96 1.1e-64 669 AZ403060 IM0170L07R Mouse
 gb_gss:AZ758880 + 876.50 1320.90 2.1e-64 652 AZ758880 IM0551B09F Mouse
 gb_est2:BF141234 + 876.50 1317.04 3.4e-64 948 BF141234 60298684F1 NCI_C
 gb_gss:AZ660560 + 872.00 1314.57 4.7e-64 620 AZ660560 IM0538P10R Mouse

seq_name: gb_gss:AZ636636

seq_documentation block: 685 bp DNA GSS 13-DEC-2000
 LOCUS AZ636636
 DEFINITION IM0495002R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0495002 R, DNA sequence.

ACCESSION AZ636636
 VERSION AZ636636
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 1 (bases 1 to 685)
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

and Wright,D., Weiss,R., Stokes,R., Tingey,A., von Niederhausern,A.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)

CONTACT: Robert B. Weiss
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177

Email: dunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0495 row: 0 column: 02
 Seq primer: CACACAGAAACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 685.
 Location/Qualifiers

1. 685
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0495002"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"

/lib_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD22nv. Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD22 (g114731114g114731114g114731114) a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 159 a 173 c 186 g 167 t

alignment_scores:

Quality: 1039.00 Length: 228
Ratio: 4.833 Gaps: 0
Percent Similarity: 94.298 Percent Identity: 84.211

alignment block:

US-09-171-553b-5 x AZ636636/rev ..

Align seg 1/1 to reverse of: AZ636636 from: 1 to: 685

```

213 AsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPr 229
|||||
685 AAFACACCACTTCTGCCGTTGAAAACAGGAGGACCAATGACTATCGCCC 636
229 OValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisPro 246
|||||
635 GSTACAGACCTCCGCGAGGTTAACAAAAGGCTCTGAACATTCACCCCA 586
246 hValProAsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsn 262
|||||
585 CAGTCCCGACCCGTACATATTATTAACCTCTCTCCACCTGAGAGACA 536
263 TrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHi 279
|||||
535 TGGTATACAGTCTGCGACTTAAAGATGCCCTTTTCCCTGCCGTTTCA 486
279 sProThrSerGlnProLeuPheAlaPheGluTTPArgAspProGlyThrG 296
|||||
485 CCTTAAAGAGTCAGCTCCTGTTTGTGAAATGAGGAGGCCACAGAGGCG 436
296 lYArgThrGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsn 312
|||||
435 GACACACGTCGTAACACTTGAGCTAGGCTACACAGAGGTTTCAAAAT 386
313 SerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheAr 329
|||||
385 TCCTCCACCTCTTTGACGAGGCCCTCCATCGGATCTTGGCCTTTTG 336
329 gIleGlnHisProGlnValThrLeuGlnTyrValAspAspLeuLeu 346
|||||
335 CGCTCGAAACCTCAGCTTACCTACTACATGATGATGATGATCTCTGG 286
346 euAlaGlyAlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeu 362
|||||
285 TTGTGGCGGCTCGAAGAGCTGTGTCAACAGGAGGACTGAGAGCTCTCTC 236
363 LeuGlnLeuSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnI 379
|||||
235 GCAGAACTGAGTGACTTGGGGTATCGAGTTTCGGCTAAAAAGCACAAAT 186
379 eCyArgArgGluValThrTyrLeuGlyTyrSerLeuArgLysGlyGln 396
|||||
185 CTGTCAAACTGAGTGAACCTACCTGGGGTATACCTCCGAGGGGTAAAA 136
396 rGTrPLeuThrGluAlaArgLysLysThrValValGlnIleProAlaPro 412
|||||
135 GATGCTCAGACAGAGGCCGGAAGAGACTGTATGATGATGCCATGCCCA 86
413 ThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysAr 429
|||||
85 ACTACCCACGAGGCGAGTGAAGTTCGTGGGAGACTGCTGCTTTGTAG 36
429 gLeuTrpIleProGlyPheAlaThrLeuAlaAla 440
|||||
35 ACTCTGATTCACAGGCTTTCACAACTCAGACGA 2
seq_name: gb_gss: BH090616
seq_documentation_block:
LOCUS BH090616 791 bp DNA GSS 18-JUL-2001
DEFINITION RPCI-24-356L4.TJ RPCI-24 Mus musculus genomic clone RPCI-24-356L4,
DNA sequence.
ACCESSION BH090616
VERSION BH090616.1 GI:14910521

```

```

KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akintore,B., Levins,M.,
Tsengaye,G., Geer,K.K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P., and Fraser,C.M.
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 791)
JOURNAL Mouse BAC End Sequences from Library RPCI-24
COMMENT
TITLE Unpublished (1999)
JOURNAL Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 356 Row: L Column: 4
Seq primer: SP6
Class: BAC ends.
FEATURES
source location/Qualifiers
1..791
/oranism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-356L4"
/clone_id="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTRABAC1; Site-1: BamHI; Site-2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 216 a 201 c 188 g 186 t
ORIGIN
alignment_scores:
Quality: 1011.50 Length: 257
Ratio: 4.341 Gaps: 2
Percent Similarity: 90.661 Percent Identity: 73.541
alignment_block:
US-09-171-553b-5 x BH090616 ..
Align seg 1/1 to: BH090616 from: 1 to: 791
118 valLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnVa 134
|||||
25 GTCTTAGCCCTCAGCTGAGGAGAAATACCATGATGATGAA...GGCAT 71
134 lYsPProAspGln.....AspIleGlnSerTrpLeuGlnGlnPheProG 149
|||||
72 AAAGCAACAGAGAGGCTGTAAGAACTGAGTGAAGCTGCTTCCCTTA 121
149 lAlaTTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPro 165
|||||
122 GAGCATGGGACAGAGACTGGAGCATGGAGATGCGTGCAGAGTCCCT 171
166 GlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTy 182
|||||
172 GTTTTGTGGGAGCTGAGACAGATGCACCCCTATAGAGAGTGCACATA 221
182 rProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgL 199
|||||
222 TCCATGAGCCGCTGAGCAGAGATGATGATGAGGCCACATATTCAGAGAC 271

```

```

199 euilleglmglylleuvalprovalglserprothrpsnthrpro 215
   ||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 TATTAACAATTAGTATTGTTGTAACCGTGCACCGTCAACCGTAATCTCT 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 leuileuprovalarglyserproglythrpsnthrargprovalglms 232
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 TTTCTCCCGGTAAAGAACCTGGCAGAGTGAATACCGACCGATCCAA 371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 pleuargluvalasnlysarvalglinspillehispthrvalproa 249
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 CCGAAGAGAGGTCAATAGAGGGTCCAGATATTCACCACTGTTCCGA 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 snpProThrAsnLeuLeuSerAlaLeuProProGluInArgAsnTrpThr 265
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 ATCCCTATACCTCCTCAGCTCAGTCCCTCCAGAGGAAATGATACACA 471
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 ValLeuAspLeuLeuAspAlaPhePheCysLeuArgLeuHisProThse 282
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
472 GTACTTGATTTAAAGATGCTTTCTTTGCTGGAAGTTACATCCCTCCAG 521
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 rglInProleuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrg 299
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
522 CCAGCCCATATTGCGCTTTGAGTGGAGAGATCCTGACACTGGACAAATGG 571
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 lylGluLeuThrTrpThrArgLeuProGluInlyPheLysAsnSerProThr 315
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
572 GACAAATTGACCTGGACACGGTTACCCAGAGAGTTCAAAAACCTCCGCCAC 621
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 llePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHis 332
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
622 CTTTGTGATGAGGCTCTACATGAGATTTAGCTCTCTCTAGTCGAA 671
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 sProGluValThrLeuLeuGlnTrpValAspAspLeuLeuAlaGlyA 349
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
672 CTCACAGGTAAACCTGCTGCATATGTTGATGACTTCTCTGCTGC.A 720
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 lathrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuLeu 365
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
721 CCACAGAGAGAGATGTTGGCAGGGGACCAAGAGATGCTAGTAGACTG 770
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 SerAspLeuGlyThrArgAla 372
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
771 GGTGAGTTGGTTACCGGGCC 791
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: gb_est2:BI107577

seq_documentation_block:

LOCUS BI107577 799 bp mRNA EST 26-JUN-2001
DEFINITION 602894023F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5039044 5',
mRNA sequence.

ACCESSION BI107577

VERSION BI107577.1 GI:14558470

KEYWORDS EST.

SOURCE

house mouse,
Mus musculus

ORGANISM

REFERENCE 1 (bases 1 to 799)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL1107 row: n column: 05
High quality sequence stop: 743.

FEATURES

source

Location/Qualifiers
1..799
/organism="Mus musculus"
/strain="C57BL/6J (f1er1)"
/db_xref="taxon:10090"
/clone="IMAGE:5039044"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 211 a 215 c 218 g 154 t 1 others

ORIGIN

alignment_scores:

Quality: 1004.00 Length: 263
Ratio: 4.236 Gaps: 5
Percent Similarity: 90.114 Percent Identity: 75.285

alignment_block:

US-09-171-553B-5 x BI107577 ..

Align seg 1/1 to: BI107577 from: 1 to: 799

```

507 ArgArgProValAlaTyrLeuSerLysLysLeuAspProValAlaSerGI 523
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20 CGGGCGCCCGTAGCGCTAGTTATCAAGAACTGGACGCTGTGGCCAGTGG 69
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
523 yTrpProValCysLeuLysAlaIleAlaIleAlaIleValLysA 540
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
70 ATGGCCCTCCTCGCTCGGAGGATAGCAGCCAGCGCTGTGTAGTAAG 119
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 spAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProHis 556
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 ATGTGTACAAACTGACATATGCGCCAGATGTTACTATAGTGGCCACAC 169
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
557 AlaLeuGlnAsnIleValArgGlnProProAspArgTrpMetThrAsnAl 573
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 TCTCTGAGAGCATCATCAGACCAACCGACCGCTGTGATGACCAACGC 219
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
573 ArgMetThrHisTyrGlnSerLeuLeuThrGluArgValThrPhea 590
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
220 CCGATAGACGCATCACAGACCTATTGCTGACAGACGAGTAAGTTTG 269
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 lAProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAsp 606
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 CACCCCGACCATTTCTAACCCCGCTCTACTACTCT...GAGGCTGAC 316
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
607 GluProValThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyA 623
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 GAGGCCCTGCACATAAGTGTGAAGAAATACTGGCAGAGACAGCTGGAAT 366
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
623 lArgGlyAspLeuThrAspIleProLeuThrGlyGluValLeuThrTrp 640
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 CCGGACACCTCACAGACCACTTGGCCAGG...GCCATGACTGGT 413
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 heThrAspGlySerSerTyrValValGluGlyLysArgMetAlaGlyAla 656
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
414 TCACAGGAGAGAGAGCTTTGTGTAGAAAGTAAAGCGAAGCGCTGGCGCT 463
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
657 AlaValValAspGlyThr.ArgThrIleThrPalasSerSerLeuProGlu 673
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 GCAGTAGTAGTGAAGACAGCGCTGTCAATATGGCCACAGCTCCGAGAG 513
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
673 lThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 689
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
514 GTACATAGCTCAAAAGCGGAACTATAGCATTAATGACGCTTAAGG 563
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
690 leuAlaGluGlySerIleAsnIleTyrThrAspSerArgTyrAlaPh 706
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

|||||.....|
564 CTGGAGAGAGAGGCTCTTAATGTCTATACCGACAGCCTGTACGCTTT 613
706 eAlaThrAlaHisValHisGlyAlaIleTyrLysGlnArgLysLeuLeuT 723
|||||.....|
614 TCCACGCGCTCATGTTCACGGAGCAATATACCGACAGCCTGTACGCTTGA 663
723 hSerAlaGlyArgGluIleLysAsn LysGluGluIleLysSerLeuLeu 739
|||||.....|
664 CCGTCGACGGCAAGATATCAAAATACACGACAGAAATTCACGCTTAT 713
714 AGAAGCGCTTTCATTTGCCCGTAGGCGATATATCATTCATTCGCCAGCA 763
756 hISGlnLysAlaLysAspLeuIleSerArgly 766
|||||.....|
764 CACCGAAGGACCTGCGCCCGCTTGAAGAGCA 796
seq_name: gb_gss:BH056750

```

```

seq_documentation_block:
LOCUS      BH056750      864 bp      DNA      GSS      18-JUL-2001
DEFINITION RPCI-24-94C16.TVB RPCI-24 Mus musculus genomic clone RPCI-24-94C16,
DNA sequence.
ACCESSION  BH056750
VERSION    BH056750.1 GI:14864050
KEYWORDS   GSS.
SOURCE      house mouse.
ORGANISM   Mus musculus.

```

```

REFERENCE
AUTHORS    Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akiref,B., Levins,M.,
            Russell,D., de Jong,P. and Fraser,C.M.
            1 (bases 1 to 864)
TITLE      Mouse BAC End Sequences from Library RPCI-24
JOURNAL    Unpublished (1999)
COMMENT    Other GSSs: RPCI-24-94C16.TJ RPCI-24-94C16.TV
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208

```

```

FEATURES
Source      Email: szhae@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cno.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 94 row: C column: 16
            Seq primer: T7
            Class: BAC ends.
            Location/Qualifiers
                1..864

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-94C16"
/clone_id="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTRABAC1; site.1: BamHI; site.2: BamHI;
RPCI-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      237 a      221 c      209 g      197 t
ORIGIN

```

alignment_scores:

Quality: 999.50 Length: 295
Ratio: 4.063 Gaps: 2
Percent Similarity: 83.390 Percent Identity: 70.508

alignment_block:

US-09-171-553b-5 x BH056750 ..

Align seg 1/1 to: BH056750 from: 1 to: 864

```

418 ValArgLupheLeuGlyThrAlaGlyPheCysArgLeuThrPleProG 434
|||||.....|
1 GTACGTAGTTCGAGACGCGAGCTTTGTAGACTCGATACCAG 50
434 yPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlyLysG 451
|||||.....|
51 GTCCACAAATGTAGCGCACCTCTGTACCCACACCTAAAGAAAGATCC 100
451 LyrPheSerThrAlaProGluHisGlnLysAlaPheAspAlaIleLys 467
|||||.....|
101 CGTTCACCTGAGACAGAAACACCAAAAGACTTTAAACACATTAAGGCT 150
468 AlaLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPh 484
|||||.....|
151 GCTCTGCTTGAGCGCCAGGCTTTGGCCCTCATATTGTACATAAGCTTT 200
484 eThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThrG 501
|||||.....|
201 CACCATATACGTTGATGAGCCAGCTGAGTAGCCGAGAGCTCTGACATC 250
501 LThrLeuGlyProTyrArgArgProValAlaLysLeuSerLysLysLeu 517
|||||.....|
251 AGACTCTGGGGCCGTGGAAGAGCTAGTGGCTACCTGTCTAAATAATTT 300
518 AspProValAlaSerGlyTyrProValCysLeuLysAlaIleAlaLys 534
|||||.....|
301 GACCAAGTTGCCATGATGAGCCCTCGCTTAAGCCATTCCTGAGT 350
534 LAlaIleLeuValLysAspAlaAspLysLeuThrLysGlnAsnLett 551
|||||.....|
351 AGCCCTACTGTTGAAGATACCTGCAAGCTTACTGAGTCAACAGTGA 400
551 hValIleAlaProHisAlaLeuGluLysnLysValAlaArgLysProAsp 567
|||||.....|
401 GTGTAATACCTCCCATGCTTAAGAGTATGTATGATGACAGCCCTGAC 450
568 ArgTyrMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuThr 584
|||||.....|
451 CATGGATGACAAATGCGCGATGACCCACTACACAGACCTGTGTCTAA 500
584 rGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 601
|||||.....|
501 TGACCGAGTGACTTTGCCCTGCTGCAATCCAAATCTACACACCTTC 550
601 euProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLeu 617
|||||.....|
551 TCCCAAGAGGATGACTCTACCTGGGTACCACAATGTGCTGATATCTG 600
618 lIleGluGluThrGlyValArgLysAspLeuThrAspLleProLeuThrG 634
|||||.....|
601 GATCAAGAAATGATGACGAGAAAGATCTGACTAC..... 636
634 yGluValLeuThrTrp.....PheThrAspGlySers 645
|||||.....|
637 .....CAAACTGGCCAGGTGTGCTGACAGGTACACAGATGTGAGCA 679
645 eTyrValValGluGlyLysArgMetAlaGlyAlaAlaValLysGly 661
|||||.....|
680 GCTTGTGTGTGAGGAGGAGAAATAGACGACGAGGAGGTGTGACAGA 729
662 ThrArgThrIleTyrPheAlaSerSerLeuProGluGlyThrSerAlaGln 678
|||||.....|
730 AAACATGTAATTTGGCGAAGTAACTTCCAGAAAGAACTTCCGCTCAGAA 779

```

678 salagileuMetalaurhnglnAlaLeuArgLeuAlaGlnGlyIys 695
 |||||
 780 GGCTGAGCTCCTGAGCTAAAGCTCCGCGCAAAAGTAGGC 828
 695 erlleAsnIleTyrThrSperArgTyrAlaPhe 706
 |||||
 829 A.ATCGAATATTTACCTGACAGCAGCATGCTTTT 863
 seq_name: gb_gss:AZ980516

seq_documentation_block:

LOCUS AZ980516 682 bp DNA GSS 27-APR-2001
 DEFINITION 2M0257006R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0257006 R, DNA sequence.

ACCESSION AZ980516
 VERSION AZ980516.1 GI:13851743

KEYWORDS

SOURCE

ORGANISM

house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

Dunn, D., Royagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

TITLE

JOURNAL

COMMENT

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0257 row: 0 column: 06
 Seq primer: CACACAGGAAACAGCATGACAC
 Class: plasmid ends
 High quality sequence stop: 682.
 Location/Qualifiers

FEATURES

source

1..682
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0257006"
 /clone_1ib="Mouse 10kb plasmid UUGC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv: Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (G147321149b|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT

143 a 171 c 189 g 179 t

alignment_scores:

Quality: 998.00 Length: 227
 Ratio: 4.752 Gaps: 1
 Percent Similarity: 92.511 Percent Identity: 81.938

alignment_block:

US-09-171-553b-5 x AZ980516/rev

Align seg 1/1 to reverse of: AZ980516 from: 1 to: 682

179 ValArgGlnTyr.ProLeuSerArgGlnAlaArgGlnGlyIleTyrPro 195
 |||||
 681 GTAGAGCAATATACCCCATGAGCAAGAGAGTAGGAGGCAATCCGGCC 632
 195 IsValGlnArgLeuIleGlnGlnGlyIleLeuValProValGlnSerPro 211
 |||||
 631 ATATCCAGAGCTTGTATGACCAAGAGAGTTTATGTCCTGTCAATCC 582
 212 TTPAsnThrProLeuProValArgLysProGlyThrAsnAspTyrAr 228
 |||||
 581 TGGATATACACCACTTCTGCGGTCGAAACCGAGGACCAATGACTATG 532
 228 gProValGlnAspLeuArgGlnValAsnLysArgValGlnAspIleHisP 245
 |||||
 531 CCGGTCGCAAGACCTCCGGGAAGTTAACAAAGAGGTCCTGGACATTCAC 482
 245 roThrValProAsnProTyrAsnLeuSerAlaLeuProGluArg 261
 |||||
 481 CCACAGTCCCGAACCCTACATTTATTAAGCTTCTCCACCCGAGAGA 432
 262 AsnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArg 278
 |||||
 431 ACATGATACAGCTCTGAGCTTAAGATGCTTCTTTCCTGGCTGGCT 382
 278 uHisProThrSerGlnProLeuPheAlaPheGlnTyrArgAspProGlyT 295
 |||||
 381 GCACCTTAGAGTACAGCTCTGTTGCTTGAATGAGAACCAACCGAGAG 332
 295 hngIyArGhrGlnGlnLeuThrTrpThrArgLeuProGlnGlyPheLys 311
 |||||
 331 GCGGACGAGCTGTCACTAACCTGACCTGACCTGACCTGACCTGACCT 282
 312 AsnSerProThrIlePheAspGlnAlaLeuHisArgAspLeuAlaAsn 328
 |||||
 281 AATTCCTCCACCTGTTTACAGAGAGCTCCATTCGGATCTGCGCTTT 232
 231 TCGCGCTCGAAACCTCAGCTTACCTACTACAGTATGATGATGATCT 182
 345 euleuAlaGlnAlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeu 361
 |||||
 181 TGGTCCGCGACCTCGAAGAGAGCTGTACACAGGAGCACTAGAGGCTC 132
 362 LeuLeuGlnLeuSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaG 378
 |||||
 131 CTCACAGAACTGAGTACTTGGGGTATCGAGTTTGGCTAAAGAGGCACA 82
 378 nIleCysArgArgGlnValThrTyrLeuGlnGlyTyrSerLeuArgGly 395
 |||||
 81 AATTGTCTCAAACTGAGGTAACTTACCTGGGATATACCTCGAGGGGGA 32
 395 InArgTrpLeuThrGlnAlaArgLysLys 404
 |||||
 31 AAAGATGCTCACAGAACCGCGGAAGAG 3

seq_name: gb_est2:BI152889

seq_documentation_block:

LOCUS BI152889 813 bp mRNA EST 05-JUL-2001
 DEFINITION 602918048F1 NC1_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5068643 5',
 mRNA sequence.

ACCESSION B1152889
 VERSION B1152889.1 GI:14612890
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 813)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILN at: <http://image.llnl.gov>
 Plate: L14M1184 row: 0 column: 12
 High quality sequence stop: 734.
 Location/Qualifiers
 1..813
 /organism="Mus musculus"
 /strain="CZECR II (feral)"
 /db_xref="taxon:10090"
 /clone="IMAGE:5068643"
 /clone_lib="NCI CGAP Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary.
 Stem cell origin."
 /lab_host="DH10B"
 /note="Organ: lung; Vector: PCMV-SPORE6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
 BASE COUNT 225 a 218 c 205 g 165 t
 ORIGIN
 alignment_scores:
 Quality: 993.00 Length: 261
 Ratio: 4.208 Gaps: 4
 Percent Similarity: 90.421 Percent Identity: 76.628
 alignment_block:
 US-09-171-553b-5 x B1152889 ..
 Align seg 1/1 to: B1152889 from: 1 to: 813
 150 AATPAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProG1 166
 |||||||
 2 GCGTGGCAGACGAGCGGAGTGGGATGGCAAAAGTG.GTCCCCCTGT 50
 |||||||
 166 nValIleGlnLeuLysAlaSerAlaThrProValArgGlnIlePro 183
 |||||||
 51 GGTATTTGAACCTTAAGTCCGGGCGCCCTTAATGAGGGGTCCGAAATATC 100
 |||||||
 183 rGleSerArgGluAlaArgGlnIleThrPro.HisValGlnArgLe 199
 |||||||
 101 CCAATGACAGAGAGAGCTCAAGGCTATACGCCCTCCAAATCAACAAC 150
 |||||||
 199 uIleGlnGlnIleLeuValProValGlnSerProThrPasnThrPro 216
 |||||||
 151 GCTCAACAAGGATTTTGTCCATGCAATCCCTTGGAAACACTCTC 200
 |||||||
 216 euleProValArgLysProGlyThrAsnSplyrArgProValGlnasp 232
 |||||||
 201 TACTTCCAGTAATAAAACCAAGGACCAAGGACTATCTCCAGTACAGGAC 250
 |||||||
 232 LeuArgGluValAsnLysArgValGlnAspIleHisProThrValProas 249
 |||||||
 251 CTTAGAGAGTCAACAAGAGAGTTCAGAGCACTACACCCCAAGGTGCCAA 300
 |||||||

249 nProTyraSnlLeuLeuSerAlaLeuProProGluArGasnTrpTyThrV 266
 |||||||
 301 TCCTTATAACCTCCTCAGACACTTGGCACCCTGGTCGACATGTGTACAG 350
 |||||||
 266 alleuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSer 282
 |||||||
 351 TCCTGATTCACAAAGACGCTTTTCTGTTGAGTTTACACCCCAACAC 400
 |||||||
 283 GlnProLeuPheAlaPheGlnTrpArgAspProGlyTyThrGlyArgThrG 299
 |||||||
 401 CAGCCCTGTGCTTCCTTCCATGATGCGAGACTCCGACAGTGGACACCCG 450
 |||||||
 299 yGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrI 316
 |||||||
 451 ACAGCTCACATGGACGAGGCTGCTCAGGATTCAGAACTGCGCCACTT 500
 |||||||
 316 lPheAspGluAlaLeuHisArgAspLeuAlaAsnThrArgIleGln.Hi 332
 |||||||
 501 TGTTCGATGAAGCCCTACACCGAGATCTGCTTTTCCGAGCCAAATA 550
 |||||||
 332 sProGlnValThrLeuLeuGlnTrpValAspAspLeuLeuAlaGlyA 349
 |||||||
 551 CCCACAGCTGACTCTCTGCAATATGTAGATGACTGCTCCTAGCTGCAG 600
 |||||||
 349 lAThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuGlnLeu 365
 |||||||
 601 AAACACGCGGAGACTGTAAATGGGACCAAAACCTCTGGGCGAGTT 650
 |||||||
 366 Ser..AspLeuGlyTyraArgAlaSerAla.LysLysAlaGlnIleCysAr 381
 |||||||
 651 AGGTAAAGCCCTGGGTATCGGGCTCTGCTTAAAGGCTCAGTATAGCCA 700
 |||||||
 381 gATGAGluValThrTyraLeuGlyTyraSerLeuArgGlyGlnArgTrpL 398
 |||||||
 701 GATAGAGTGTACTACTAGATATGTCTTGACAGATGACCAAGTGTAG 750
 |||||||
 398 eutThrGluAlaArgLysLysThrVal 406
 |||||||
 751 TCACAGAGCCAGACAAAGACTGTA 776
 |||||||
 seq_name: gb_gss:A2627566
 seq_documentation_block:
 LOCUS A2627566 690 bp DNA GSS 13-DEC-2000
 DEFINITION IM0469L20F Mouse 10kb plasmid UNGC1M library Mus musculus genomic
 clone UNGC1M0469L20 F, DNA sequence.
 ACCESSION A2627566
 VERSION A2627566.1 GI:11749756
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 690)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0469 row: L column: 20
 Seq primer: CTTGTATAAACGACGGCCAGT

Class: plasmid ends
High quality sequence stop: 690.
Location/Qualifiers
1. 690

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U08C1M0469L20"
/clone.lib="Mouse 10kb plasmid U08C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 182 a 192 c 165 g 151 t
ORIGIN

alignment_scores:
Quality: 988.00 Length: 228
Ratio: 4.682 Gaps: 1
Percent Similarity: 92.544 Percent Identity: 81.579

alignment_block:
US-09-171-553B-5 x AZ627566 ..

Align seg 1/1 to: AZ627566 from: 1 to: 690

243 ILEHSPHITHTVALPROASNPROTYRASNLEULEUSERALALEUPROPR 259
|||||
7 ATACACCCACAGTCCCAATCTTATACCTCCACAGCCTGGCCACC 56
259 OGLUARGASNTPTTTHRVALLLEUASPLEULYSAPALAPHEPCYSL 276
| |||||
57 TGGTGGACATGGGTACACAGTCTGATCTCAAGATGCTTTTCTGTT 106
276 EUATRLHISPROTHRSERGINPROLEUPHEALAPHEGLTPARGASP 292
|||||
107 TGAGGTACACCCCAACAGCAGCCTTGTTCGCTTTCGATGCGGAGAC 156
293 PROGLYTHIRGLYARGTHTGILEUHRHTPHRARGLEUPROGLINGI 309
| |||||
157 TCCGAGATGGACAGCCGAGACGCTCAATGAGAGGAGCTGCTCAAGGG 206
309 YPHELYSANSERPROTHRIEAPHSPLUALALEUHISARGASPLEUA 326
|||||
207 ATTCAAGAACTGCGCCACTTTGTCGATGAAGCCCTACAGGAGTCTTG 256
326 LAASPHARGILLEGHISPROGLINVALTHIRLEULEGLTYRVALASP 342
| |||||
257 CTCCTTCCGAGCCCAATTAACCAACAGGACTCTTTCGCAATATGATGAT 306
343 ASPLEULEUVALAGLYALATHTIRYSLINSPCYSLLEUGLUGLYTHIRLY 359
| |||||
307 GATCGCTGCTTACGCTGACAGAAACAGCAGAGAGCTGTGAATTTGGACCCA 356
359 SALALEULEUENGLIULEUSERTASPLEUGLYTYRARGALASERIALALYSTL 376

357 AAACCTCTCGGCGAGATTAGTAGCTGGGTTATCGGCGCTGCTGCTAAA 406
376 YSALAGINILIECYSAARGGLUVALTHIRYLEUGLYTYRSELEUARG 392
| |||||
407 AGGCTGATATGACATGAGTGAAGTCACTACCTAGAGATATGCTTGAGA 456
393 GLYGLYGLINARGTPTPLEUHTHGLVALARGLYSLYSTRVALVALGINIL 409
| |||||
457 GATGGACACAGGTGGCTCACAGAAAGCCAGAAACACACTGTTATGCAGAT 506
409 EPICOLAPROTHRTHALALYSGLNVALARGLUPELEUGLYTHRALAG 426
| |||||
507 CCGGCGCCCAACACACTGCTGCCAGGTAAAGAGATTCTTGGGACGCCG 556
426 LYPHECYSAARGLEUTPTILEPROGLYPHEALATHIRLEUALAIPROLEU 442
557 GGTTCACAGACTCTGTGATTCGCGGATTCACACAGCGCAGCTCCCTTA 606
443 TYRPROLEUHTIRYSLYGLYGLY. PHESERTPTALAPROGLUHTISG 459
| |||||
607 TATCCTACTTACCAAGAGAAAGGGGATTCACCTGACACAGACATC 656
459 INLYSALAPHEASPALALIELYSLYSALALEU 469
| |||||
657 AGCTAGCCTTTGAACATCTCAAAAAGGCACTG 688

seq_name: gb_gss:AZ717731

seq_documentation_block:

LOCUS AZ717731 817 bp DNA 24-JAN-2001
DEFINITION RPCI-24-15511.TJ RPCI-24 Mus musculus genomic clone RPCI-24-15511,
DNA sequence.

ACCESSION AZ717731.1 GI:12456936

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other_GSSs: RPCI-24-15511.TV

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1998)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@ligr.org

Library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/ordering.htm). BAC end

plate: 155 row: I column: 1

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. 817

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-15511"

/clone.lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBA1; site_1: BamHI; site_2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

630 CATTCTCTGCCCCGGCCACCAAGAGAGAGAATCTGGTGGCCAGGGCA 699

US-09-171

US-09-171

Align seg 1/1 to: A2715956 from: 1 to: 789

```

527 CysleuysAlaIleAlaValAlaIleValLeuValLysAspAlaAspIy 543
|||||
5 TGTCTGAAGAGCTATTGACAGACAGTACCGCTGTGATTAAAGATGGAGCAA 54
543 sleuthrleuGlyGlnAsnIleThrValIleAlaProHisAlaLeuGln 560
|||||
55 ATTGACAAATGGGACAAAGAGTACGTGTGTGGCCCTCATGCTTAGAAA 104
560 snIleValArgIleProProAspArgTpmethrAsnAlaArgmethr 576
|||||
105 GTATGCTGGCTGACGACCTGACAGATGATGACAAATGGCCGATGCA 154
577 HisThrGlnSerLeuLeuLeuThrGlnArgValThrPheAlaProAla 593
|||||
155 CACTATCAGAGCTGCTGCTTAATGATGATGATGATGATGATGATGATG 204
593 AlaLeuAsnProAlaThrLeuLeuProGluThrAspGluProValT 610
|||||
205 CATCTCAACCCAGCTACCGCTCTCCCTTACCAATGATTCCTCCAG 254
610 hHisAspCysHisGlnLeuLeuIleGluThrGlyValArgLysAsp 626
|||||
255 TACATCAATGATGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 304
627 LeuThrAspIleProLeuThrGlyGluValLeuThrThrPheThrAspG 643
|||||
305 CTGACTGACCAACCTGGCTGGA...GCTCCCACTGCTGCTGCTGCTGCT 351
643 ySerSerThrValValGluGlyLysArgMetAlaGlyAlaAlaValAla 660
|||||
352 CAGCAGATTCTGATAGAGGAAAGCGAAAGGCTGAGCTGGGTGTAG 401
660 spGlyThrArgThrIleThrPalaSerSerLeuProGluGlyThrSera 676
|||||
402 ATGGGAAAAAGGTAATTGGGCAACGCTTGTCTGAAGGAAACGTCGGA 451
677 GlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGlu 693
|||||
452 CAAAAGCTGAACTTATACGCTTATTCAGCCCTCCGAGAGGCTTAAAG 501
693 yLysSerIleAsnIleThrThrAspSerArgTyrAlaPheAlaThrAla 710
|||||
502 TAAGATCATTAACATCTACATGACAGCCGCTATGCTTTCCTACAGC 551
710 IValHisGlyAlaIleThrLysGlnArgGlyLeuLeuThrSeraIle 726
|||||
552 ACATCCATGGGGCCATCTACAGGCAAGGGCTATTGACCTGGCTGCT 601
727 ArgGluIleLysAsnLysGluGluLeuLeuSerLeuLeuGluAlaLeu 743
|||||
602 AAAGACATTTAAACAAAGAAATTTCTGCGCTTGTGGAAGCCATACA 651
743 sLeuProLysArgLeuAlaIleIleHisCysProGlyHisGlnLysAla 760
|||||
652 TGCACCTAGAAGAGTACCATCATCAGTCCGCGGCGCAAGAGAGAG 701
760 yAspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLysGln 776
|||||
702 AAGACTTGTGGCCAGGCGCAACCGAATGACAGACTACGCAAAACAA 751
777 AlaAlaGln 779
|||||
752 GTTGTCTCAA 760

```

seq_name: gb_est2:BF784826

seq_documentation_block: 975 bp mRNA

LOCUS BF784826 975 bp mRNA EST 12-JAN-2001

DEFINITION 602111174F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4239307

ACCESSION BF784826

VERSION BF784826.1 GI:12089862
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 975)
 AUTHORS NIH-MGC
 TITLE NIH-MGC http://mgi.nhl.gov/
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rmail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM9852 row: c column: 20
 High quality sequence stop: 701.
 FEATURES
 source
 1..975
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:4239307"
 /clone_host="NCI CGAP Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site: 1; Note: Site: 2; Sall; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. 1"

BASE COUNT 257 a 281 c 252 g 185 t
 ORIGIN

alignment_scores:
 Quality: 967.00 Length: 349
 Ratio: 3.748 Gaps: 8
 Percent Similarity: 73.926 Percent Identity: 60.172

alignment_block:
 US-09-171-553B-5 x BF784826 ..

Align seg 1/1 to: BF784826 from: 1 to: 975

```

157 MetGlyLeuAlaLysGlnValProProGlnValIleGlnLeuLysAla 173
|||||
3 GTGGGAGATGGCAAACTGGTCCCGCTGCTGATGATGATGATGATGATG 52
173 rAlaThrProValSerValArgGlnIleThrProLeuSerArgGluAlaArg 190
|||||
53 GGCACCCCTATAGAGGGTCCGACAAATATCCATAGCAGAGAACTCAAG 102
190 IugIlyIleThrProHisValGlnArgLeuIleGlnGlnGlyIleLeuVal 206
|||||
103 AGGTAATACGCCCAATTAACAACCTGCTCAACAAAGGAGATTGTGCT 152
207 ProValGlnSerProThrPalaThrProLeuLeuProValArgLysPro 223
|||||
153 CCATGCAAAATCCCTTGGAAACATCTCTACCTCCAGTAAACAAACCAAG 202
223 yThrAsnAspTyrArgProValGlnAspLeuArgIleValAsnLysArgV 240
|||||
203 GGACCAAGACTACCGCTCCAGTACAGAGACTTACAGAACTCAACAGAG 252
240 aGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSerAla 256
|||||
253 TTCAGGACATACACCCCGCTGCGCAAAATCTTATTAACCTCCCTGACG 302
257 LeuProProGluArgAsnIleThrThrValLeuAspLeuLysAspAlaPh 273
|||||

```

```

303 TTGGCAGCTGGTGGACATGTACACAGTCTGGATCTCAAGAGCGTTT 352
273 ePhcysLeuArgLeuHisProthrsSerGlnProleuPheAlaPheGlnT 290
353 TTTCTCTTTGAGGTACACCCCAACAGCCCTTTGCTTGCCTTCGAAAT 402
290 rPArspProglYthrGlyArgThrGlyGlnLeuThrTrpThrArgLeu 306
403 GGGAGACTCCGAGAGTGCACAAAGCCGACACTCATGACGAGAGGCTG 452
307 ProGlnGlyPheLysAsnSerProthrllePheAspGlnAlaLeuHisAr 323
453 CCTCAGGAGATTCAGAACTCGCCACTTTGTCGATGAAGCCCTACACCG 502
323 gAspleuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnT 340
503 AGATCTTGCTCTTTTCCGAGCAATACCCACAGGACTCTTCTGCAAT 552
340 yTValAspAspleuLeuAlaGlyAlaThrLysGlnAspCysLeuGln 356
553 ATGTAGATGACCTGCTCTAGCTGCAGAAACACGAGAGACTGTGAAT 602
357 GlyThrLysAlaLeuLeuLeuLeuLeuSerAspLeuGlyTyrArgAlaSe 373
603 GGGACCCAAACTCTGGGCGAGTTAGTTAAAGCTGGGATACGGGCTTC 652
373 rAlaLysLysAlaGlnIleCysArgArgGlnValThrTyrLeuGlyTyrS 390
653 TGCTAA.AAGGCTCACTTATGCCAGATAGAAATGACCTACCTAGATATG 701
390 eTLeuArgGly GlyLlnArgTrpLeuThrGlnAlaArgLysLysThrVa 406
702 TCTTGAGCAGATGGGCGACGGGTG.....TCCAGAAACAGAAACAG 745
406 lValGlnIleProAlaProThrThrAlaLysGlnValArgGlnPheLeu 423
746 TGTTCAGATCCCGAGCCACATGTCCGACGA.....GAGTCTCTGG 789
423 lYThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAla 439
790 GGCCCGCGGGT...TGCGACGTG.....TGCGACGTG..... 809
440 AlaProLeuTyrProLeuThrLysGlnLysGlyLysPheSerTrpAlaPr 456
810 ATCCGATTGGCCA.....TGGGCTTC 832
456 oGlnHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAlaP 473
833 TTGTTCCCTACCAAGAGGATCCGCGCAACACACTGTTGACCAAA 882
473 rAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 489
883 GCGGGGCGCGGGGCCCA...GTAACTCTCCCT..... 911
490 GluArgGlyGlyValAlaArgGlyValLeuThrGlnThrLeuGly 504
912 CCGGAAGAAGGGGCGCCAGCGAGACACAAAGAGGCGC 956

seq_name: gb-gss:BH106844

```

```

seg documentation block:
LOCUS      BH106844      842 bp      DNA
DEFINITION RPECI-24-352A17.TVB RPECI-24 Mus musculus genomic clone
ACCESSION  BH106844
VERSION    BH106844.1 GI:14937281
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 842)
AUTHORS   Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,

```

```

FEATURES
    source          1..842
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone.lib="RPECI-24-352A17"
                    /sex="Male"
                    /cell_type="Spleen/Brain"
                    /note="Vector: pTABBAC1; Site_1: BamHI; Site_2: BamHI;
                    RPECI-24 Mouse BAC library produced by Pieter de Jong. The
                    library was cloned in the pTABBAC1 cloning vector at the
                    BamHI sites using MboI partially digested male C57BL/6J
                    DNA."
BASE COUNT      221 a      228 c      200 g      193 t
ORIGIN
alignment_scores:
    Quality:      966.50      Length:      267
    Ratio:        4.221      Gaps:        1
    Percent Similarity: 85.768      Percent Identity: 70.412
alignment_block:
US-09-171-553B-5 x BH106844
Align seg 1/1 to: BH106844 from: 1 to: 842
373 SerAlaLysLysAlaGlnIleCysArgArgGlnValThrTyrLeuGlnTyr 389
18 ACCACTAAGAAAGCTCAGTATGCCATTTGCACTTGACCTAGCAAGATA 67
389 rSerLeuArgGlyGlyGlnArgTrpLeuThrGlnAlaArgLysLysThrV 406
68 CTCCTCTCGAAGAAAGCAGTGGTTCACCGAAGCTCGAAGAAAGACATG 117
406 AlValGlnIleProAlaProThrThrAlaLysGlnValArgGlnPheLeu 422
118 TGATGCAAGATCCCAACCCCAACCACTGTGACAAAGTACGTGAGTTTCTG 167
423 GlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAl 439
168 GGAACGCGCAGGCTTTGTAGACTGTGAATACAGGGTCCACAAATGTTGGC 217
439 AlaProLeuTyrProLeuThrLysGlnLysGlyLysPheSerTrpAlaP 456
218 GGCACCTGTGACCACTCACTAAGAAAGTCCCGCTTCACTGACGACG 267
456 rGlnHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAla 472
268 AGGAGACGCCAAAGAGCTTTAAACAGCATAAAGGCTGCTGCTTGCAGCC 317
473 ProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAs 489

```

```

|||||
318 CAGCGCTTGGCCCTGCCAGATTGACTTAACCTTTACCCATATACGTGA 367
489 pgluarqlysglyvalaalaarglyvalleuthrlnthrleuglypro 506
|||||
368 TGACCCAGCTGGAGTAGCCCGCAGAGTCTGTACATCACTCTGGGGCC 417
506 rpatagatpovaalalatyrlleuserlyslsleuappprovalalaser 522
|||||
418 GGAAGAGGCCAGTGGCTTACCTGTCTAAAAAATTTGACCCAGTGGCCAG 467
523 glytrpovaalysleuysalalealavalaleuvally 539
|||||
468 GAATGGCCCTCTCTCTTAAAAAGCATGCTGTATACCTCTCTGTATA 517
539 sasplalaspysleuthrleuglylnasnllethvallealaproh 556
|||||
518 AGATACCTGGCAAGTTACTGTAGGTGACATGTGAGTTTATACCTCC 567
556 isalaleuqluasnllevalarglnproprobspargtmetthrasn 572
|||||
568 ATGCTTAAGATGATTGTATGACAGCCCTGACCATGTGATGACAAAT 617
573 Alarqmethrhrsttyrglnserleuileuthrlnthrph 589
|||||
618 GCCCGATGACCCACCTACACAGCCTGTGTATATGACGAGTACTTT 667
589 ealaprovalaalaaleuasnpovalathrleuileuproglyluthra 606
|||||
668 TGCCCTACCTGCCATCCACATCTACACCCCTCTCCACAGAGGATG 717
606 spgluprovalthrhisaspeyhslnleuileuglyluthrly 622
|||||
718 CCTTACCTGGGTACCAATGCTGTATTTCTGGCTGAAGAAGATGGA 767
623 Valatqlyasplleuthrasplleproleuthrlylualleuthrtr 639
|||||
768 TCGAAGAAAGATCTGACTGAC.....CGAAGCTTG 796
639 p 639
797 g 797

seq_name: gb_gss:A2715272
seq_documentation_block: 755 bp DNA 24-JAN-2001
LOCUS A2715272
DEFINITION RPI-24-154N5.tj RPI-24 Mus musculus genomic clone RPI-24-154N5,
DNA sequence.
ACCESSION A2715272
VERSION A2715272.1 GI:12451823
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 755)
AUTHORS Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
Isegaay,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPI-24-154N5.tj
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC

```

```

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 154 row: N column: 5
Seq primer: SP6
Class: BAC ends.
FEATURES
    source
        Location/Qualifiers
            1..755
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="RPI-24-154N5"
                /clone_1id="RPI-24"
                /sex="Male"
                /cell_type="Spleen/Brain"
                /note="Vector: pTRABAC1; Site_1: BamHI; Site_2: BamHI;
                RPI-24 Mouse BAC Library produced by Pieter de Jong. The
                library was cloned in the pTRABAC1 cloning vector at the
                BamHI sites using MhoI partially digested male C57BL/6J
                DNA."
BASE COUNT      202 a      194 c      190 g      169 t
ORIGIN
alignment_scores:
    Quality: 961.00      Length: 253
    Ratio: 4.271      Gaps: 3
    Percent Similarity: 88.933      Percent Identity: 75.099
alignment_block:
    US-09-171-553b-5 x A2715272 ..
Align seg 1/1 to: A2715272 from: 1 to: 755
273 PhephecysleuargyleuHISProthrSerglnProleuPhealaphel 289
|||||
5 TTTTCTGTTGGCT...TCACCCACAGCCAGCCCTGTTCCTTCCA 51
289 utrpargaspptrogllythrlyargthrlyglnleuthrtrpthrargl 306
|||||
52 ATGGTGAGACTCCGAGAGTGAGACAGCGGACGACATGACGAGAGGC 101
306 euproglynglypheylsasnserrprothrlietheaspqlyalaleuHIS 322
|||||
102 TGCCCTCAGGATTCAGAAACCTCGCCCACTTTGTTCATGAAGCCCTACAC 151
323 ArgaspleualiasnphearglielglnHISproglNvalthrleuengl 339
|||||
152 CGAGATCTTGCTCTTTCCAGCCAAATTAACCCAGAGTGTCTTCGCA 201
339 ntyrvalaspaspleuileuualaglyalathrlysglnaspCysleug 356
|||||
202 ATATGTAGATGACCTGCTCTAGCTGCAGAAACAGCAGAGACTGTGAA 251
356 lueglythrlyAlaleuileuHISleuserlytyrtyrtyrtyrtyrtyr 372
|||||
252 TTGGGACCCAAACCTCTGGGAGTTAGGTAAAGCTGGGGTATTCGGGCC 301
373 Seralalysysalaglnilecysargarggluvalthrtyrleuglyty 389
|||||
302 TCTCTCTAAAAAGCTCGATTATGACGATTAAGTGAAGCTACCTAGGATA 351
389 rserleuarglylyglnargtrpleuthrlnthrlylualarglyslslythrv 406
|||||
352 TGCTTTAGAGATGAGACAGCGTGGCTGCACAAAACAGAAACAGAAACAGCTG 401
406 alvalglnileproalaprothrthrAlatysglnvalatrgluPheleu 422
|||||
402 TTATGACAGATCCGACCCCAACCACTGCTGCCGAGTAAGAGATTCTCTG 451
423 glythrAlaglypheycsargleutrpilreproglyPhealathleual 439
|||||
452 GGAACCGCGGGTTTTCAGACTCTGATTCGCCGATTTCGACACTACACATGAC 501

```

```

439 aAlaProLeuTyrProLeuThrLysGluLysGlyGlyPheSerTrpAlaP 456
|||||
502 AGCTCCCTTTGATCCACTACCAAGAGAGGGAATTCACCTGGATG 550
|||||
456 roGluHisGluLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAla 472
|||||
551 A.GAACATCAGCTAGACCTTTGAACTCTCAAAAGGCACTCTCCAGGCT 600
|||||
473 ProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValas 489
|||||
601 CCGCATTTGGCCCTGCCAGATTAAACAAACCTTTCACCCATATGATGA 650
|||||
489 pGluArgGlyGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProt 506
|||||
651 TGAAGAATGAGCTGCGAAGGAGGCGCTTACCCAGGTTTGGAGACCAT 700
|||||
506 rPATATGProValAlaTyrLeuSerLysLysLeuAspProValAlaase 522
|||||
701 GGAAGCGCCCGGTAGCTACTTAT..AAGAACTACAGCTGTGGCAG 747
|||||
522 rGlyTyr 524
|||||
748 TGGATGG 754
|||||

```

seq_name: gb_AZ827524

seq_documentation_block: 700 bp DNA
 LOCUS AZ827524
 DEFINITION 2M0104E02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 accession AZ827524
 keywords AZ827524.1 GI:12997432
 version
 source house mouse.
 organism Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Dunn, D., Moyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)

JOURNAL
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177

FEATURES
 source
 1..700
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0104E02"
 /clone_id="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
 /note="Vector: pMD42hy: Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (914732114[gb|AF129072.1]) a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed mouse DNA was selected for ampicillin resistance.

alignment_block:
 US-09-171-553b-5 x AZ827524

alignment_scores:
 Quality: 960.50 Length: 233
 Ratio: 4.488 Gaps: 1
 Percent Similarity: 91.845 Percent Identity: 76.824

Align seg 1/1 to: AZ827524 from: 1 to: 700

```

416 LysGlnValArgGluPheLeuThrAlaGlyPheCysArgLeuTrpI 432
|||||
5 CGGACGATGAGTGGAGTCTTCTGGGACTGCTGCTTTGAGACTCTGAT 54
|||||
432 eProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 449
|||||
55 TCCAGGCTTTGCACCTTACGACGACCTCTATATCTTGTGATTAAGAG 104
|||||
449 yGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIle 465
|||||
105 GGGTTCCTTCGAGTGGAAAGAAAGAACACAGAGAGCTTTGAGGCTATC 154
|||||
466 LysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThr 482
|||||
155 AAGTGTCTCTATATGACTGCCCCCGGCTAGCATTACCAAGACTTGACTAA 204
|||||
482 sProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 499
|||||
205 GCCTTTCGCTCTATATGTGAGACGAGAGCGGGGTATACCGAGGAGCT 254
|||||
499 eutThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLys 515
|||||
255 TGACACAAGCACTGGACCTGGAGAGAGACCTGTAGCTATTGTCAAAA 304
|||||
516 LysLeuAspProValAlaSerGlyTyrProValCysLeuLysAlaIle 532
|||||
305 AAATTAGATCTCTGTTGCTAGTGGATGCCCCACATGTCTGAAGCTATTGC 354
|||||
532 AlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyLys 549
|||||
355 AGCATACCTCTGTTGATCAAAAGATGCTGACAAATTGACATGGAGACAC 404
|||||
549 snIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnPro 565
|||||
405 AGGTGACTGTTGTGGCCCTCATCTAGAAATATCGTCCGACACCA 454
|||||
566 ProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeu 582
|||||
455 CCGACAGATGATGACAAATGCGGATGACACACTATAGAGCCGCT 504
|||||
582 uLeuThrGluArgValThrPheAlaProAlaAlaLeuAspProAla 599
|||||
505 GCTAAATGAGGCTGTAACCTTTGGCCCGCTGCATCTCAACCCAGCTA 554
|||||
599 hrLeuLeuProGluGluThrAspGluProValThrHisAspCysHisGln 615
|||||

```



```

300 .....GlnLeuThrTrpThrArgLeuProG 308
753 CGAGAAAAGAGTTCAGAAAACATGAGGCTGACCAAGAC..... 792
308 lnglYpHeuLysAsnSerProThrIlePheAspGluAlaLeuHis...Arg 323
793 .....GAGACGAGAGTCAGAAATCAGACGCTGCTGTGAGACCACGA 834
324 AspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTty 340
835 GAGATCAGAAAATGAAATGTCGGCCACCAAGTCCGGAATAATVAGAAA 884
340 rValAspAspLeu.....LeuValIleGlnValThrL 351
885 CTTAGTNTTACCTGCAACAAAAAAGCTCTTGCT.....C 925
351 ysgLAspCysLeu...GluGlyThrLysAlaLeuLeuGluLeuSer 366
926 AAGATCAATGTGTTACGATTAACAAAAGGGGTCTAATACAG..... 969
367 AspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgI 383
970 .....AAGCTGAGTCTCTAGAAAGAGTCC 992
383 uValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrG 400
993 GGTGGCCTAACCTG.....TTGAAGC 1012
400 lValArgLysLysThrValValGlnIleProAlaProThrThrAlaLys 416
1013 ACCTGAAAAAGAGACTGTGTTCACTCCACCGACGACGATGCAAAA 1062
417 GlnValArgGluPheLeuGlyThrIleGlyPheCysArgLeuTrpIlePr 433
1063 CAGCTAAAAAGTCTCGGCGACTCGGGCTTTTGCAGATTGTGATTCC 1112
433 oGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysG 450
1113 AGTTTGCTTAAT.....TAAAGAGATAAC 1138
450 lylGlyPheSerTrpAlaProGluHisGln.....LysAlaPheAspAla 464
1139 AGCCCTGTATAGAAAATATAAAACAACTTGATGGCTTGATGCC 1188
465 IleLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValTh 481
1188 ATTAAGACTGCCCTGATGTATCCCACTATGGGACTCTAGATGTGAC 1238
481 lLysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyV 498
1239 T.....GAGACAAAGCTATTGTCAAGAGAG 1264
498 alleuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSe 514
1265 TTCTTACCTAGAGATTGGACCTAGAAAAGACCTGTGCTACTTGT.. 1312
514 lLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaI 531
1313 ..AAAAATTAGACCTGCTGCTGTAGATGGCTGCTGTGTGAC.... 1357
531 leAlaIleValAla..IleLeuValLysAspAlaAspLysLeuThrLeuG 547
1358 .....ATAGTGGCTTCTGTGTCAGAGACGACAGATAATTGACTGTGAG 1399
547 yGlnAsnIle.....ThrValIleAlaProHis..... 556
1400 ACAAAACTTGCCACATGCTCTAAAAAGTGTGTTCAAGCCCAACCGCA 1449
557 .....AlaLeuGluAsnIleValArg..GlnProProAspArgT 569
1450 TGGCTGACTAATGCTCTTGAAAACATTATCCACACGCTCCCTGACGAT 1499
569 rPmetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGlu 585
1500 GG.....ACACATTATCAGACCTTTTGTCTACTGAA 1531
586 ArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeuLeuP 602
1532 CGAGTGAACCTTCCTCCCTCCCTGCTGCTCAAT.....CTCACTACTGCG 1575
602 roGluGluThrAspGluProValThrHisAspCysHisGlnLeuLeuIle 618
1576 CTGAGACTTCCCT.....ACTCATATTGTGCTGACATTCAGCA 1616
619 GluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrGly.. 634
1617 GAAGAACTCATACTCCAAATGATCTAAAGATAGATCAGCTGGCCT 1666
635 GluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGly 651
1667 GAGAGTTTGAGCTGTGACAGGATGCGATGACCTGAGGTTGAGGGTAA 1716
651 sArgMetAlaGly...Ala..AlaValValAspGlyThrArgThrIleTr 666
1717 GCGGAAGCGGAGACAGTGCAGTGTGACAGCAAGAAAGATGATCTA 1766
666 rAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMet 683
1767 GCCCAGACGCTCCCTTAAGGAGCTTCAAGCCGTAAGCTTAACCTTGC 1816
683 lAlaThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyr 699
1817 CTTTAATATCAACCT..... 1830
700 ThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAla..IleT 716
1831 .....CTAATATAT 1838
716 yrlYsglnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLys 732
1839 ACAGACAAAGAGAGCTGTGACATCTGCA..... 1867
733 GluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAl 749
1867 ..... 1867
749 allelleHisCysProGlyHisGln..LysAlaLysAspLeuIleSerArg 765
1868 .....GAAAAGACCTTAAGTGTGCTGTGCTAAA 1895
766 GlYAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAs 782
1896 AAAAATCAGATGGCAATCTAACCCGCCAGCATCTCTAAAGACATGAT 1945
782 nLeuLeuProIleIleGluThrProLysAlaProGluProArgArgGlnT 799
1946 CCTG..... 1949
799 yrlThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerG 815
1950 ..ACAGTCTAAAGACTATCAAGTTATAGACAAATTA..... 1983
815 uThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProH 832
1983 ..... 1983
832 isLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeu 848
1983 ..... 1983
849 GlYThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValle 865
1984 .....ACAGTGTAAAGAAACCTGTATGACATGAT 2015
865 uArgLeuProGlyValAlaAspSerValValLysHisCysValProCysG 882

```



```
2016 AGAAACTGAAGGACACGAATATAGCTCTCA..... 2049
882 InLeuValAsnAlaAsnProSerArgIleProGlyLysArgLeuArg 898
2050 .....TGA 2052
899 GlySerHisProGlyAlaHisTrpGluValAspPheThrGluValLysPr 915
2053 GAAGCAGACCTGACATCTGACAAAGATAGACTTCACTGAAAGAAAAAT 2102
915 oAlaLysTrpGlyAsnLysTrpLeuLeuValPheValAspPheSerG 932
2103 GTGATG.....AATACCTTCTG.....TTTGTGACGTTCTCA 2140
932 LTTTPValGluAlaTrpProThrLysLysGluThrSerThrValValAla 948
2141 GATGGATAAAGCTTTCCTTATAAACTGAGACTACTGACATAGTCATC 2190
949 LysLysIleLeuGluGluIlePheProArgPheGlyIleProLysVal 965
2191 AAGAAGATTATTAATAAT...TTTCCAAGTTCGAGTGCCTCAAGCAAT 2237
965 eGlySerSpasnglyProAlaPheValAlaGlnValSerGlnGlyLeu 982
2238 AGTGTCAATATAGTCTGCTGCTTGTGCTCCAGTAAGTCAGGCTGTGG 2287
982 IaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTrpArgProG 998
2288 CCAGTATTATAGAGTCAATACAAAGATG.....TACAGACCTCA 2328
998 nSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeu 1015
2329 GAACCTCAGACAGATAGATAATATAGATAAAACTCT.AAACAGACTTGA 2377
1015 hLysIleuThrThrGluThrGlyIleAsnAsp.TirMelAlaLeuLeuPr 1031
2378 CAAATTATATCTAGAGACTGGC...ACAGACTTACTTGGTGGCCCTTCC 2424
1031 oPheValLeuPheArgValAlaArgAsnThrProGlyGlnPheGlyLeuThr 1048
2425 CCTGGCCCTATTAAACTGAGATACCTCCCTTGATAGGCTTTTACTC 2474
1048 roTyrgIleLeuLeuTyrgIlyProProLeuAlaGluIleAlaPhe 1064
2475 TTTTAAAGATCCTTTATGGGGCTCTATGCCCATCCTGTCTTAATAAT 2524
1065 AlaHisSerAlaAspValLeuLeu.SerGlnProLeuPheSerArgLeu. 1080
2525 GTGTTTAA...CCTATGTGTATATAATATGATCTGTATACATAATTAT 2571
1081 LysAlaLeuGluTrpValArgGlnArgAla.....\TrpLysGln.L 1094
2572 AAGCCTTACAGGTGTCAGAAAGAGAGTGGTCACAACAGCTACAGTG 2621
1094 euArgGluAlaTrpSerGlyLysPheGlnValProHisArgPheGln 1110
2622 AACGAGCTGGGTACCCCAAGGACATCTTACCAGTTCCAGCCAGA..... 2665
1111 ValGlyAspSerValTyValArgArgHisArgAlaGlnLysLeuGluTh 1127
2666 .....GATCGATCTGCTGACACCTGCTCATCTGAGACCTCAAGCC 2709
1127 rArgTrpLysGlyProTyrgLeuValLeuLeuThrThrPro 1140
2710 TCACAGAAAGGCTCCCTGCTAGTCTGTACTAATCCG 2749
```

OM of: US-09-171-553b-6 to: EST: * out_format: pfs
Date: Feb 24, 2002 8:31 AM
About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: -DEV=xlh
-MODEL=frame+g2n.model
-O=/cgn2.1/USP10_spool/US09171553/runat_22022002.122825.11137/app_query.fasta.1.3312
-DB=EST -QFMT=fastap -SUFFIX=feb22std.rst -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPEXT=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FEAPOP=6.000 -FEAEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DEL0P=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09171553.ecgnl.1.9008 -NCPV=6 -ICPU=3 -LONGLOG
-DEV_TIMECUT=120 -WARN_TIMECUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:
Query: US-09-171-553b-6
Query length: 656
Database: EST: *
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 3515.700000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est1:AM657531	894.00	1242.03	5.1e-60	549	1 AM657531 110922 MARC BOV Bos t
gb_gss:AZ970955	881.50	1221.03	7.1e-59	754	1 AZ970955 2M0244H16F Mouse 10kb
gb_gss:BM059027	865.50	1197.63	1.5e-57	882	1 BM059027 RPCI-24-326G3.TJ RPCI-
gb_gss:BH122287	849.50	1176.62	2.2e-56	774	1 BH122287 RPCI-24-288P13.TJ RPCI-
gb_gss:AZ351242	820.00	1136.40	3.9e-54	701	1 AZ351242 1M0089109F Mouse 10kb
gb_est1:BI151181	787.00	1088.55	1.0e-51	845	1 BI151181 602916935F1 NCI CGAP
gb_gss:AZ850764	776.50	1076.55	8.4e-51	641	1 AZ850764 2M0552C18R Mouse 10kb
gb_gss:AZ813042	763.00	1058.65	8.4e-50	625	1 AZ813042 2M0808E11F Mouse 10kb
gb_est1:BI106669	763.00	1054.22	1.5e-49	921	1 BI106669 602933272F1 NCI CGAP
gb_gss:AZ356870	740.50	1026.82	5.3e-48	614	1 AZ356870 1M0098809F Mouse 10kb
gb_gss:AZ376517	730.50	1012.82	3.0e-47	640	1 AZ376517 1M0130202F Mouse 10kb
gb_gss:AZ400249	729.00	1011.04	3.8e-47	591	1 AZ400249 1M0214G11R Mouse 10kb
gb_gss:AZ379933	727.50	1007.07	5.8e-47	673	1 AZ379933 1M0688E24R Mouse 10kb
gb_gss:AZ642336	721.50	1000.05	1.5e-46	624	1 AZ642336 1M0505M18F Mouse 10kb
gb_gss:AZ741412	719.00	993.11	2.4e-46	659	1 AZ741412 RPCI-24-74C13.TJ RPCI-
gb_gss:AZ747473	717.00	988.76	3.7e-46	632	1 AZ747473 1M0050K14R Mouse 10kb
gb_gss:AZ789547	697.00	965.08	1.4e-44	632	1 AZ789547 RPCI-24-73J17.TJ RPCI-
gb_gss:AZ735836	696.50	966.11	1.2e-44	565	1 AZ735836 1M0129J20F Mouse 10kb
gb_gss:AZ499035	693.00	961.77	2.1e-44	534	1 AZ499035 1M0336P05R Mouse 10kb
gb_gss:AZ417785	687.50	950.36	9.0e-44	787	1 AZ417785 602901065F1 NCI CGAP
gb_gss:AZ417785	683.50	947.66	7.1e-43	833	1 AZ417785 MY1.001169 Mouse 9-day
gb_est1:BE569478	676.00	933.76	7.6e-43	895	1 BE569478 601328494F1 NCI CGAP
gb_gss:AZ620796	673.00	928.89	1.2e-42	650	1 AZ620796 1M0453112R Mouse 10kb
gb_gss:AZ620796	670.00	928.67	1.5e-42	592	1 AZ620796 2M0131E01F Mouse 10kb
gb_gss:AZ636232	669.50	926.37	2.0e-42	605	1 AZ636232 2M0191D01R Mouse 10kb
gb_gss:AZ416306	668.00	911.99	1.2e-41	680	1 BF159914 601767582F1 NCI CGAP
gb_est1:BF159914	659.00	911.39	1.3e-41	724	1 BF159914 601767582F1 NCI CGAP
gb_gss:BM054319	658.50	912.10	1.2e-41	625	1 BM054319 RPCI-24-398E11.TV RPCI-
gb_gss:AZ336017	649.00	898.63	1.1e-40	969	1 BF181106 601805164F1 NCI CGAP
gb_est1:BF181106	648.50	898.40	7.1e-41	609	1 BF181106 601805164F1 NCI CGAP
gb_gss:AZ647666	646.00	888.28	2.1e-41	679	1 AZ647666 1M0544J12F Mouse 10kb
gb_gss:AZ647666	626.50	865.97	4.5e-39	728	1 BI106591 602893174F1 NCI CGAP
gb_est1:BI106591	625.50	864.74	5.3e-39	716	1 BI106591 602921838F1 NCI CGAP
gb_gss:AZ89131	624.00	854.18	5.7e-39	656	1 AZ89131 2M0272E22F Mouse 10kb
gb_gss:BI082444	622.50	859.41	1.0e-38	750	1 BI082444 602877669F1 NCI CGAP
gb_gss:AZ349077	619.00	859.26	1.1e-38	493	1 AZ349077 1M0085K31R Mouse 10kb
gb_gss:AZ960032	618.00	858.29	1.0e-38	644	1 AZ960032 2M0227K32R Mouse 10kb
gb_gss:AZ397096	615.50	851.35	3.0e-38	675	1 AZ397096 1M0161J32R Mouse 10kb

gb_gss:AZ996293 - 609.50 844.67 6.9e-38 566 1 AZ996293 2M0282L19F Mouse
gb_est1:BF786434 - 606.00 836.44 2.0e-37 801 1 BF786434 602112966F1 NCI CGAP
gb_gss:AZ887402 - 605.00 838.24 1.6e-37 575 1 AZ887402 RPCI-24-176N10.TJ
gb_gss:BM055839 - 602.00 833.32 3.0e-37 620 1 BM055839 RPCI-24-93P16.TVB
gb_gss:AZ018624 - 601.00 831.70 3.7e-37 635 1 AZ018624 RPCI-23-382C22.TV

seq_name: gb_est1:AM657531

seq_documentation block: 549 bp mRNA EST 25-APR-2001
LOCUS AM657531
DEFINITION 110922 MARC BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AM657531
VERSION AM657531.1 GI:7423429
KEYWORDS EST.
SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 549)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Cases,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-Mckinn,C.G.,
Pettea,G., Holt,I., Karaycheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.

REFERENCE
1 (bases 1 to 549)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Cases,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-Mckinn,C.G.,
Pettea,G., Holt,I., Karaycheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
Sequence evaluation of four pooled tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smiththermal.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCATGACAT
BACKWARD: GTTTCAGCAGCAGCAGC
Plate: 94 row: E column: 12
Seq primer: ATTAGTACATATAG.

FEATURES
source 1..549
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC BOV"
/cissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: xbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

BASE COUNT
145 a 144 c 121 g 139 t
ORIGIN
alignment_scores:
Quality: 894.00 Length: 183
Ratio: 5.109 Gaps: 1
Percent Similarity: 95.628 Percent Identity: 90.164

alignment_block:
US-09-171-553b-6 x AM657531 ..
Align seg 1/1 to: AM657531 from: 1 to: 549
289 serProglyValProvalysThrGlyGlnArgLeuPheSerLeuIleG1 305
|||||
3 TCCCCCGGTGTCCTGTTAAACAGCAGCAGCAGCTCTTCATCA 52
|||||
305 nGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerc 322
|||||

```

53 GGGAGCTTTCCAAAGCATCACTCAACGACCCGATGCCAGTTCTT 102
322 YSTPLPLeuSerSerSerGlyProProTyrTyrGlucMetAlaLys 338
103 GTTGGCTTGTCTATCTCAGGGCCCTTATATAGAGGAATGGCTAGA 152
339 GLUArgLysPheAsnValThrlYsgLuhSArgAsnGlcysThrPrpL 355
153 GAAGGAATTCATCTGACCAAGAGCATAGAAATCAATGTCACATGGG 202
355 YSerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysI 372
203 GTCCCAAAATAGCTCACCCTCAGTCAAGTTCCGGGAGGGAGACATGA 252
372 LGLYsAlaProProSerHisGlnHisLeuLysCysTyrSerThrValAl 388
253 TGGAAAGACTCCCTCCATCCACCAACCTTTGCATATGACTATGCTT 302
389 TYRGlulAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTr 405
303 TATGAGCAGCGCTTAGAAAATCATATTTAGTACCTGGTTATACCTGAG 352
405 PTPAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheA 422
353 GTGGCAGTGCATCTGAGGTAACTCCTGCTTTCTACACAGCTTTTA 402
422 snGLSerLysAspPheCysValMetValGlnIleValProArgValTyr 438
403 ACCAATCCAAAGATTCTGTGTGTCATGTCAGCTGCTCCCTGGGTACAC 452
439 TYRHisProGluGluValAlaValLeuAspGluTyrAspTyrArgTyrAsnArg 455
453 TACCATCTCGAGGAAGTAGTATCATGATAGTATGCTATTCGCGCTACCCG 502
455 gProLysArgGluProValSerLeuThrLeuAlaValMetLeuGly 470
503 CTCAAAAGAGAGCGCCGTGACCTTAACCTTAGCGGTTATCTAGGA 549
seq_name: gb_A2970955

```

seq_documentation_block:

LOCUS A2970955 754 bp DNA GSS 27-APR-2001
 DEFINITION 2M0244H16F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0244H16 F, DNA sequence.

ACCESSION A2970955
 VERSION A2970955.1 GI:13842182

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunne@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0244 row: H column: 16
 Seq primer: CATTGTAAAGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 754.
 Location/Qualifiers

source

```

1. 754
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0244H16"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-gold, T1-resistant, F-"
/notes="Vector: PM042nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (GI4732114) (pBlAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 226 a 160 c 150 g 218 t
ORIGIN

```

alignment_scores:

Quality:	Ratio:	Length:
881.50	4.025	244
Percent Similarity: 89.754	Percent Identity: 67.623	

alignment_block:

US-09-171-553b-6 x A2970955/rev

Align seg 1/1 to reverse of: A2970955 from: 1 to: 754

```

406 TTPAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAs 422
|||||  |||||||  |||  |||||
751 TGGGCTTGCAGAAACCGTCTCATCTCTGTGTGTCATGTCAGATTGTA 702
422 ngInserLysAspPheCysValMetValGlnIleValProArgValTyr 439
|||||  |||||||  |||||||  |||||||  |||
701 TGGTCCAAAGATTCTGTGTGTCATGTCAGCTTCTTCTACACCTCTGT 652
439 YTHisProGluGluValAlaValLeuAspGluTyrAspTyrArgTyrAsnArg 455
|||||  |||||||  |||||||  |||||||  |||
651 ATCATGATGATAGTCCTCTTTAGACAAATTTGACATCGGGTCC...CGC 605
456 gProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuG 472
|||||  |||||||  |||||||  |||||||  |||
604 TGGAGAGAGAGCCCATTAACCTTAAGTTCATTAAGATGATGAG 555
472 YTHraAlaValGlyValGlyThrGlyThrAlaAlaLeuThrGlyProG 489
|||||  |||||||  |||||||  |||||||  |||
554 AGTACGCGCTGAGTAGTACGAGAACCGCTGCTTAATTAAGACCCGCC 505
504 AATACTAATGAA.....GAATCAGCTGACAGCTATGAGATGTTGA 467
505 PLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThr 522
|||||  |||||||  |||||||  |||||||  |||
466 TCTTGAAGCTATAGACAGTCTATACCAATTTAGAGAAATCTTAACCT 417
522 erLeuSerGluValAlaLeuGlnAsnArgArgGlyLeuAspLeuLeuPhe 538
|||||  |||||||  |||||||  |||||||  |||
416 CCTGTCCGAAGTGCTGCTACAGATAGAGAGGATTAAGTATATTTTC 367

```

539 LeuargGluGlyGlyLeuGlyAlaAlaLeuGlyGluGlyCysCysPheTy 555
 366 CTTAAAGAGAGGAGACTCTGCTGCTCTAAAGAGAGATGTGTTTA 317
 555 rValAspHisSerGlyAlaAlaLeuArgAspSerMetAsnLysLeuArgLysL 572
 316 TCTTACCATTCAGAGATCAATCAAGATTCATGCGCAACTGAGAAC 267
 572 ySLeuGluArgArgArgGluArgGluAlaAspGlnGlyTyrPheGlu 588
 266 GCGTGAATACACGTAAAGAGGAAAGAGCAACAGAGATGTTGCA 217
 589 GlyTyrPheAsnArgSerProTyrPheThrPheLeuSerAlaLeuTh 605
 216 ACTGCTGTTAATAGTCCCTTGCTCACACTCTCTCTCCACCATAGC 167
 605 rGlyProLeuValValLeuLeuLeuLeuThrValGlyProCysLeuI 622
 166 AGGACTTTGATTACTCTTATCTTTGCTTCTTGGCCCTCATCC 117
 622 LeAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnLe 638
 116 TTAATAGTACTAGCTTTTATTAGAGAAAGATTAATGACATGAGGTT 67
 639 MetValLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 648
 66 ATGCTACTAGGCAACATATCGGCTCTT 37
 seq_name: qb_gss:BH059027

seq_documentation_block: 882 bp DNA 18-JUL-2001
 LOCUS BH059027
 DEFINITION RPI-24-326G3.TJ RPI-24 Mus musculus genomic clone RPI-24-326G3,
 DNA sequence.
 ACCESSION BH059027 GI:14868646
 VERSION
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 882)
 REFERENCE
 AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinfet, B., Levins, M.,
 Tesgaye, G., Geer, K., Krol, M., Shartsbeyn, A., Gebregeorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPI-24
 Unpublished (1999)
 other_GSSs: RPI-24-326G3.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@tigr.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderingframe.html>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 326 row: G column: 3
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..882
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-24-326G3"
 /clone_11b="RPI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;

RPI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamH1 sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 248 a 183 c 201 g 250 t
 ORIGIN

alignment_scores:
 Quality: 865.50 Length: 309
 Ratio: 3.652 Gaps: 6
 Percent Similarity: 76.699 Percent Identity: 58.252

alignment_block:
 US-09-171-553B-6 x BH059027 ..

Align seq 1/1 to: BH059027 from: 1 to: 882

297 GYLnArGLeuPheSerLeuLeuGlnGlyAlaPheGlnAlaLeuAsnSe 313
 2 GGGCAAGAGAGCTTCACACCTGATTAGAGAGAGCTTTATGCTTACAG 51
 313 rThAspProAspAlaThrSerSerCysTyrPheCysLeuSerSerGlyP 330
 52 CTGCGATCCAGATGCTCCATCATCTGCTGTTATTCCTATCATCAGGCC 101
 330 rOPrOTYrTYrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 346
 102 CTCCCTACTCTGAGAGGGTTCCTTTTATGAGAGTTTATTAACACAGT 151
 347 GlnHisArgAsnGlnCysTyrTyrGlySerArgAsnLysLeuThrLeu.. 362
 152 AGTCATGCC...TTATGCTCATGGGCAACAGAAAGAAATTAACCTGTA 198
 363 ...ThrGluValSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 378
 199 CTGAGATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 248
 378 eRhISGlnHisLeuCysTyrSerThrValValTyrGlnGlnAlaSerGI 394
 249 CCCACCAACACCTTATGTCGTGGAGCTTATTCGTATCCAGAGACAAATAC 298
 394 uASnGlnTyrLeuValProGlyTyrAsnArgTyrTyrPalaCysAsnThrc 411
 299 CAACATATTACCTTGTCATATCCAGTTGTTGGTGGCCCTGTAACACAG 348
 411 lYLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspHe 427
 349 GGCCTCACCCCATGTGTATCAACCTAGTGTGCTTGGACCCCTCGGTGATTC 398
 428 CysValMetValGlnIleValProArgValTyrTyrHisProGluGluVa 444
 399 TGTATCATGCTTCATCTTCTATCTGCGATGATATTATTCCTGCTCCGG 448
 444 lValLeuAspGluTyrArgTyrArgTyrAsnArgProLysArgLuprov 461
 449 CCGTGAAGATGATATATATACAGTAAG...AAGTTGAAGAGATCCAG 492
 461 aLSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyAl 477
 493 TCTCCATATCTAGTACTAGTTAAATGGGGGCGGCTTACAGA..GGAATA 541
 478 GlyThrGlyTyrAlaAlaLeuLeuThrGlyProGlnGlnLeuGlyGlu 494
 542 GGAATGGGTTTAGCAGCTTGGTTAGGGGTAGACAGAAATTCACACTT 591
 494 yLeuGlyLeuLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 511
 592 CTTAAGAGAA.....GCTATAAATGAAGATTAGCATGCTACAAA 632
 511 lUSeValSerAsnLeuGluGlnSerLeuThrSerLeuSerGluValAla 527
 633 AATCTGTGATGATAGAGAAATCTTGGAGCTTCCCTGTCAGAAAGTAGTA 682

528 LeuClInaSnArGArGcLyLeuAspLeuLeuPheLeuArgGluGlyLeu 544
 683 TTACAAAACAGAGAGGTAGATTATGTTTAAAGGAGAGGTGACT 732
 544 uCysAlaLeuLeuGluGlyCysCysPheTyrValAspHisSerGlyA 561
 733 ATGTGCAGCCTTAAGAGAGAAATGCTTCTTATGACAGACACGCGGAA 782
 561 laIleArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArg 577
 783 TAGTCAGAGATCTATGCAAAATGAGGAAAGACTGGAAGGGA... 829
 578 ArgGluArgGluAlaAspGlnGlyTrrPheGluGlyTrrPheAsnArg 594
 830 TGGTTCAGAGTGTGACTCCTAATC 855
 594 rProTrrMetThrThrLeuLeuSer 602
 856 TCCATGATGACACCTTACTTCT 880

seq_name: gb_gss:BH122287

seq_documentation_block:

LOCUS BH122287 774 bp DNA GSS 19-JUL-2001
 DEFINITION RPI-24-288P13.TJ RPI-24 Mus musculus genomic clone RPI-24-288P13
 , DNA sequence.

ACCESSION BH122287

VERSION BH122287.1 GI:14965799

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 774)
 Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinet, B., Levins, M.,
 Tesgaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPI-24

AUTHORS

TITLE Unpublished (1999)
 JOURNAL Other-GSS: RPI-24-288P13.TJ
 COMMENT Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@ligr.org

Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
 page: http://www.ligr.org/lbd/bac_ends/mouse/bac_end_intro.html
 Plate: 288 row: P column: 13
 Seq primer: SP6
 Class: BAC ends.

FEATURES

location/Qualifiers

1..774
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-24-288P13"
 /clone_lib="RPI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAcl; site_1: BamHI; site_2: BamHI;
 RPI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAcl cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT

221 a 162 c 172 g 219 t

alignment_scores:
 Quality: 849.50 Length: 262
 Ratio: 3.776 Gaps: 2
 Percent Similarity: 85.878 Percent Identity: 62.214

alignment_block:
 US-09-171-553B-6 x BH122287 ..

Align seg 1/1 to: BH122287 from: 1 to: 774

354 TrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyGlyLeu 370
 2 TTTGGGAGAAACAAAGTTGACTCTGCGACAGCTTTCAGAGAGAGGCT 51
 370 rCysIleGlyLysAlaProPheSerHisGlnHisLeuCysTyrSerThr 387
 52 TTGTTGGGCGAGGACTCAGAGTAAAGGCGACCTGTATATCAGACC 101
 387 AlaValTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsn 403
 102 AGAATCATCAGTCTAGCAAAAGTGTGATGATGATGATGATGATGAT 151
 404 ArgTrrPrrAlaCysAsnThrGlyLeuThrProCysValSerThrSer 420
 152 ACAGTGGGCGCTGCAATACCGGTCTCAGTCTGCTGCTGCTGCTGCT 201
 420 LPhasnGlnSerLysAspPheCysValMetValGlnIleValProArg 437
 202 TTTTAATATGTTCCAAAGATTTCTGATTTGGTTCACCTTATTCCTAG 251
 437 aLYrTYrHisProGluGluValValLeuAspGlnTyrAspTyrArg 453
 252 TCTGATATCATGATGATAGGCGCTTTTACCAAAATTTGAGCATGGGCT 301
 454 AsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeu 470
 302 ...CGGTGAGAGAGAGAGCCGTTACCTTACCTTGGCAGTCTTATAGG 348
 470 YLeuGlnThrAlaValAlaGlyThrGlyThrAlaAlaIleThrG 487
 349 ATTGAGAGTACGGCTGAGTACGAGACACCGCTTATTTATTTAGA 398
 487 LyrProGlnGlnLeuGluLysGlyLeuGlyLysLeuHisAlaMetThr 503
 399 CCCCCCAATAGTATGAA.....CACTACGTCGACGTATGAT 436
 504 GluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluSerLe 520
 437 GTTATCTTATGAACTATGAGACAGTCTATACCAAAATTTAGAGAAATCT 486
 520 uThrSerLeuSerGluValAlaLeuGlnAsnArgGlyLeuAspLeu 537
 487 AACCTCCCTCGCGAAGTGTGCTACCAAAATTAAGGAGATTGACTTAT 536
 537 euPheLeuArgGluGlyLeuCysAlaAlaLeuLysGluGlyCys 553
 537 TATTCCTTAAAGACGAGACTCTGCTGCCCTAAAGAGAAATTTGA 586
 554 PheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeu 570
 587 TTTTATGTTGACCATTCATGATGATTAACAAAGATTTATGCGCAAACTTG 636
 570 gLysLysLeuGluArgArgArgGluArgGluAlaAspGlnGlyTrrP 587
 637 AGAAGCGCTGATATACGTGAAGAGAAATTTAGAAACCAACAGATGT 686
 587 heGluGlyTrrPheAsnArgSerProTrrMetThrThrLeuSerAla 603
 687 TCGAAAGCTGGTTAATAGTCCCTTGGCTGACCACTCTCTCTCCAC 736
 604 LeuThrGlyProLeuValValLeuLeuLeuLeu 615

737 ATAGAGAGACCTTGATTACTTATGCTTGGCT 772

seq_name: gb_gss:A2351242

seq_documentation_block:

LOCUS A2351242 701 bp DNA GSS 29-SEP-2000
DEFINITION IM0089109F Mouse 10kb plasmid UUGCIM library Mus musculus genomic
clone UUGCIM0089109 F, DNA sequence.ACCESSION A2351242
VERSION A2351242.1 GI:10430479KEYWORDS GSS.
SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 701)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,
Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
and Wright,D., Weiss,R.TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid insertsJOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome CenterKm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0089 row: 1 column: 09Seq primer: GGTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 701.

FEATURES

Location/Qualifiers

1..701
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCIM0089109"
/clone_lib="Mouse 10kb plasmid UUGCIM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adapter oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (q114732114[gb|AF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."BASE COUNT 211 a 149 c 134 g 206 t 1 others
ORIGINalignment_scores:
Quality: 820.00 Length: 234
Ratio: 3.923 Gaps: 4
Percent Similarity: 89.316 Percent Identity: 67.949alignment_block:
US-09-171-553b-6 x A2351242/rev ..

Align seg 1/1 to reverse of: A2351242 from: 1 to: 701

416 ValSerThrSerValPheasnInSerLysAspPheCysValMetValG1 432
||||| |||||||
700 GTGCTGATGCTGCTTTTAACTAGTCCAAAGATTCTGCATTTGGTTCA 651
432 nllValProArValYrThrHisProGluGluValValLeuAspGluT 449
|||||
650 ACTTATTCCTAGACCTCGTATCATGATGATAGTACCTCTTTTAAACAA 601
449 YrAspYrArGTYrAsnArGProLysArGluProValSerLeuThrLe 465
|||||
600 TTGAGCATCGGGTC...CGCTGAGAGAAGAACCCGTTACCTTACTTT 554
465 uAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrA 482
|||||
553 GCGAGTCTCTATTAGATTGGAGTAGCGGTTGGAGTAGAGAGAACTG 504
482 lAlaLeuLeuThrGlyProGluGluGluGlyGlyGlyGlyGlyLeu 498
|||||
503 CTGCTTAATTAAAGACCCCAATACTATGAA.....GACTTA 466
499 HisAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAs 515
465 CGTCACTACTATGATGATGTTGATCTTAGAACATTAAGAACAGCTA 416
515 nLeuGluGluSerLeuThrSerLeuSerGluValValLeuGluAsnAr 532
|||||
415 ATTAGAAGAAATCTTAATCTCCCTCGAAGGCGTGTCAAGAAATGAA 366
532 rGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeu 548
|||||
365 GGGGATTAGACTTATTATTCCTTAA...GGAGACTCTGTGCTGCCCTA 319
549 LysGluGluCysCysPheYrValAspHisSerGlyAlaLeuArgAsp 565
|||||
318 AAAAAGAGTGTGTTTATGTTGACCATTCAGAGATTAATCAAGATTC 269
565 rMetAsnLysLeuArGlyLysLeuGluArGArGArGArGLeuArgGlu 582
|||||
268 TATGCCCAACTTAGAGACGCCCTAGATATACGTAAAGAGAGAGAGAA 219
582 lAspGluGlyTrpPheGluGlyTrpPheAsnArGserProTrpMetThr 598
|||||
218 GCCAACAAGATGGTTGAAAGCTGTTTATATAGTCCCTGGCTACCC 169
599 ThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLe 615
|||||
168 ACTTCCTCCTCACCACATAGCAGACGCTTATTAATCACTTATGCTTGT 119
615 uThrValGlyProCysLeuLeuAsnArPheValAlaPheValArgGlu 632
|||||
118 TACTTTGGCCCTGACATCCCTTAATAGTAGTGTATTAATTAAGAGAA 69
632 rGValSerAlaValGluLeuMetValLeuArGInGInTrGInGlyLeu 648
|||||
68 GGATTAATAGCAGTACAGGTTATGGTACTAAGCCACAATATCGGCTCTT 19

seq_name: gb_est2:B1151181

seq_documentation_block:

LOCUS B1151181 845 bp mRNA EST 05-JUL-2001
DEFINITION 602916935F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5067273 5',
mRNA sequence.ACCESSION B1151181
VERSION B1151181.1 GI:14611182KEYWORDS EST.
SOURCE house mouse.ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

349  TTTCATCATGA..TTTGTGCATGATCCAGCTGTATACCCGGGTATTT 397
349  yHfAsPProGluGluValLeuAspGluYrAspYrArgYrAsnArg 455
398  ATACACCTGCATCCAGTTTA.....GAAGAACCTATGCTGGCCGGCGG 441
456  ProlYsArgGluProValSerLeuThrLeuAlaValMetLeuGluLeuG 472
442  TCAAAAAGAGAACCAATTACTTTAACCTGGCTGCTCATTTACGGGAATAG 491
472  yTHrAlaValAlGlyValGlyThrGlyThAlaAlaLeuIleThGlyProG 489
492  TATGCAAGTATAGAGATGGGAGCGGAGGTGTACGTTGTATGAAGGAAGAC 541
489  InGluLeuGluYrGlyLeuGlyGluLeuHsAlaAlaMetThrGluAsp 505
542  AG.....GGAAATTCAGTCTTTGAGGAGATGCTGCAATGAAAGAC 579
506  LeuAlaGAlaLeuYrGluSerValSerAsnLeuGluGluSerLeuThrSe 522
580  CTAGCGGCAATGAGAGAACTCCATTCAGACGCTTTGAAAACCTTTGACAC 629
522  rLeuSerGluValValLeuGlnAsnArgArgGly..LeuAspLeuLeuPhe 538
630  CCTGCTCAGGTAGTATTACAGAAAGAGAGAGGAGGTTCGATTTGGTGTTC 679
539  LeuArgGluGlyGlyLeuCySAlaAlaLeuYrGlu..GluCySAsPheT 555
680  CTACAGGAAGAGGAGACTGTGTGCTGCCCTTAAGAAAGAGTGTGCTTTCT 729
555  yValAsPHisSerGlyAlaIleArg..AspSerMetAsnYrLeuArgly 571
730  ATGCAGATCATCACAGGAATAGTTGAAGACTCTATCCGAACCTGAGAG 779
571  sLys...LeuGluArgArgArgGluArgGluAlaAspGlnGlyTTPP 587
780  AAGGATTCGAGCGAGGAGACCGGAGAACCGGATGCTCAACGGGGTGGTG 829
587  heGluGly 589
830  AGTCGGG 837

seq_name: gb_gss:A2850764

seq_documentation block:
LOCUS      A2850764          641 bp      DNA          GSS          21-FEB-2001
DEFINITION  ZM0152C18R Mouse 10kb plasmid U06C1M library Mus musculus genomic
ACCESSION  A2850764
VERSION    A2850764.1 GI:13036088
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 641)
AUTHORS    Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
JOURNAL    Contact: Robert B. Weiss
COMMENT    University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00

```


musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMDA2 (g1473214|g149129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 191 a 128 c 120 g 186 t
ORIGIN

alignment_scores:
Quality: 763.50 Length: 207
Ratio: 4.083 Gaps: 2
Percent Similarity: 90.338 Percent Identity: 69.565

alignment_block:
US-09-171-553b-6 x AZ813042/rev

Align seg 1/1 to reverse of: AZ813042 from: 1 to: 625

```

407 AAlAGSAsnThrGlyLeuThrProCysValSerThrSerValIleAsnGly 423
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
607 GCTTCACATTAACGGCTCACCCTGCTGCTATGCTGTTTAAATAG 558
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
423 nSerLysAspPheCysValMetValGlnIleValProArgValIleTyrH 440
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 TTCTAAAGATTTCTGCATTTGGTTCACGCTTATCCGACCTCTGTATC 508
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
440 lSProGluValIleuAspGluTyrAspTyrArgTyrAsnArgPro 456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
507 ATGATGATAGCTCTTTTACATTAATTTGACATCGGCTC...CGCTGG 461
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
457 LysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyTh 473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 AAAAGAGAACCCGTTACTTACCTTGGCAGTTCTTATAGATGGAGCT 411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
473 rAlaValGlyValGlyThrGlyThrAlaAlaIleThrGlyProGln 490
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 AGCAGCTGGAGTAGTAGAGAACCGCTGCTTATTAATTAAGACCCCAAT 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
490 lLeuGluTyrGlyLeuGlyLeuLeuHisAlaIleMetThrGluAspLeu 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 ACTATGAA.....GACTACGTGACGCTATGATGATTAATGCTT 323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
507 ArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSer 523
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 AGAAGCTATAGACAGCTCTATTAACCAATTAAGAAATCTTAATCTCCCT 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
523 uSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 GTCGGAAGTGGTCTGCACAAATAGAGGGGATTAATTAATCTCCCTTA 223
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
540 rGluGluGlyLeuGluCysAlaAlaLeuLysGluLysCysPheTyrVal 556
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 AAGAGAGAGAGACTCTGTGCTCTTAAAGAAAGATGTTGTTTATGTT 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 AspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLe 573
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 GACCATTCAGGAGTATCAAGATTTCTATGGCTAAACTTAGAGAACGCTT 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
573 uGluArgArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyT 590
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 AGATTAAGTAAAGAAAGAGAGAGCCACACAGATGGTTCCAAAGCT 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
590 rPheAsnArgSerProTyrPheThrThrLeuLeuSerAlaLeuThrGly 606
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 GATTAAATAGTACCTCCCTGCTCACACACTCTCTCTCCACTATAGCAGGA 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 CTTTAAATTAACCTTAATGCTT 2
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: gb_est2:B1106669
seq_documentation_block: 921 bp mRNA EST 26-JUN-2001
LOCUS B1106669
DEFINITION 60288322F1 NC1_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5038336 5',
mRNA sequence.
ACCESSION B1106669
VERSION B1106669.1 GI:14557562
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 921)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM11105 row: p column: 17
High quality sequence stop: 805.
Location/Qualifiers
1..921
/organism="Mus musculus"
/strain="C57BL/6J (f1)"
/db_xref="taxon:10090"
/clone_image="5038336"
/clone_id="NC1_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary."
/stem_cell_origin=""
/lab_host="DH10B"
/note="Organ: Lung; Vector: pCMV-SPORT6; Site_1: Salt.
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 249 a 211 c 229 g 232 t
ORIGIN

alignment_scores:
Quality: 763.00 Length: 304
Ratio: 3.206 Gaps: 10
Percent Similarity: 78.289 Percent Identity: 54.276

alignment_block:
US-09-171-553b-6 x B1106669

Align seg 1/1 to: B1106669 from: 1 to: 921

```

298 GlnAgluPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerTh 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 CAGAGAGATGTTTAACTAGTAGAGAGACCTTTATGCTTATACAGAC 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 rAspProAspAlaThrSerSerCysTyrPheCysLeuSerSerGlyProp 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

[illegible]

LOCUS	AZ356870	644 bp	DNA	GSS	02-OCT-2000
DEFINITION	M0099F09F09E Mouse 10kb plasmid UGCM library Mus musculus genomic clone UGCM0099F09 F, DNA sequence.				
ACCESSION	AZ356870				
VERSION	AZ356870.1	GI:10470556			
SOURCE	GSS.				
ORGANISM	house mouse. Mus musculus				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus-1 (bases 1 to 644)				
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Petersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.				
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunne@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0098 row: F column: 09 Seq primer: CGTTGTAAACGACGGCCAGT Class: plasmid insert High quality sequence stop: 644. Location/Qualifiers 1..644				

```

BASE COUNT      185 a      136 c      133 g      150 t
ORIGIN

alignment_scores:
  Quality:      740.50      Length:      215
  Ratio:        3.897      Gaps:        2
  Percent Similarity: 88.372  Percent Identity: 66.977

```

```
seq_name: gb_gss:AZ356870
seq_documentation_block:
```

```

alignment block:
US-09-171-553B-6 x AZ356870/rev ..
Align seg 1/1 to reverse of: AZ356870 from: 1 to: 644

```

```

390 GluAlaIaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTr 406
      :::::::::::::::::::::
639 CAGCTAGTAAGAGTGTGAGTATCTAGTCCCTCCGACACAGACATATG 590
      :::::::::::::::::::::
406 pAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnG 423
      :::::::::::::::::::::
589 GGGTGGCAATACCGGTCTCAGCTCTGTGTCTATGCTTTAATA 540
      :::::::::::::::::::::
423 InsertAspPheCysValMetValGlnIleValProArgValLysTrpTr 439
      :::::::::::::::::::::
539 GTTCCAAAATTTCTTCATTTGTGTCAGCTTATCTTCCAGACCTGTAT 490
      :::::::::::::::::::::
440 HisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPr 456
      :::::::::::::::::::::
489 CATTGATGATGATGCTCTTTTAGCAAAATTTGACATCGGCTC...CGCTG 443
      :::::::::::::::::::::
456 OlyAspGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyT 473
      :::::::::::::::::::::
442 GAGAGAGAACCCGCTTACCTTACTTGGCAGTTTA.TTAGATTTGGAG 394
      :::::::::::::::::::::
473 hrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGln 489
      :::::::::::::::::::::
393 TACGGCTGGAGTAGGTACAGAGAACCGCTTATTAAGAACCCCA 344
      :::::::::::::::::::::
490 GlnLeuGluLysGlyLeuGlyLeuHisAlaAlaMetThrGluAspLe 506
      :::::::::::::::::::::
343 TACTATGAAAAA.....CTAGCTGACGATATGATGATCT 306
      :::::::::::::::::::::
506 uArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerL 523
      :::::::::::::::::::::
305 TGAACATATAGAACACTATTAACCAATTAAGAAATCTTAACTTCC 256
      :::::::::::::::::::::
523 euserGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuLeu 539
      :::::::::::::::::::::
255 TGTCCTCAATGTGCTCTACAGAAATAGAGGGGATTAAGCTTATTAATCT 206
      :::::::::::::::::::::
540 ArgGluGlyGlyLeuGlyAlaAlaLeuLysGluGlyCysPheTyrVa 556
      :::::::::::::::::::::
205 AAGAAGAGAGACTGTGCTCTCCCTAAAGAAATGTGTTTATG 156
      :::::::::::::::::::::
556 IAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysL 573
      :::::::::::::::::::::
155 TGACCATTCACGAGTAATCAAGATTTCAVGGCCAACTTAGAGAACCC 106
      :::::::::::::::::::::
573 euGluArgArgArgArgGluArgGluAlaAspGluGlyTrpPheGluGly 589
      :::::::::::::::::::::
105 TGAATATACGTAAAGAGAAAGAGAACCAAGATGTTCCAAAGC 56
      :::::::::::::::::::::
590 TrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeu 604
      :::::::::::::::::::::
55 TGGTTTAATAGTCCCTTGGCTCCACACTCTCTCCACCATTA 11

```

seq_name: gb_AZ376517

seq_documentation_block:

LOCUS AZ376517 610 bp DNA GSS 02-OCT-2000

DEFINITION IM0130J20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION AZ376517.1 GI:10490217

VERSION AZ376517.1 GI:10490217

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 610)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

and Wright,D., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL Plasmid Inserts

unpublished (2000)

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0130 row: J column: 20
Seq primer: CGTTGTAAGAACGACGGCCAGCT
Class: plasmid ends
High quality sequence stop: 610.

FEATURES

source

1. 610

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0130J20"

/sex="Male"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD42ny. Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt-end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pMD42 (g11473211419b1AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adaptor mouse DNA was annealed to

adaptor vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 177 a 131 c 120 g 182 t

ORIGIN

alignment_scores:

Quality: 730.50 Length: 204

Ratio: 3.970 Gaps: 2

Percent Similarity: 90.196 Percent Identity: 67.157

alignment_block:

US-09-171-553B-6 x AZ376517/rev ..

Align seg 1/1 to reverse of: AZ376517 from: 1 to: 610

```

431 ValGlnIleValProArgValTyrTyrHisProGluGluValValLeuAs 447
      :::::::::::::::::::::
608 GTTCAGCTTATTCCTAGACTCTGTATCATGATGATGATGCTCTTTTACA 559
      :::::::::::::::::::::
447 pGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuT 464
      :::::::::::::::::::::
558 CAGATTGACATCGGCTC...CGCTGAGAAAGAGCCCGTTACCTTAA 512
      :::::::::::::::::::::
464 hrLeuAlaValMetLeuGlyLeuGlyThrAlaValAlaGlyThrGly 480
      :::::::::::::::::::::
511 CTTTGGAGTTCTATAGATTTAGAGATGACGGCTGATAGTACAGCA 462
      :::::::::::::::::::::
481 ThrAlaAlaLeuIleThrGlyProGlnGlnLeuGlyGlyLeuGlyG 497
      :::::::::::::::::::::
461 ACCGCTGCCCTTAATTAAGACCCCAATACTATGAA.....CA 424
      :::::::::::::::::::::
497 uLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValS 514
      :::::::::::::::::::::

```

```

423 ACTAGTCGACGATGATGTTGATCTTGAACCTATAGACAGCTCTATA 374
514 eAsnLeuGlusSerLeuThrSerLeuSerGluValValLeuGlnAsn 530
373 CCAAAATAGAGAAATCTTAACTTCCCTGCGAAGTGTGCTACAGAAAT 324
531 ATGATGlyLeuAspLeuThrLeuArgGluGlyGlyLeuCysAlaI 547
323 AGAAGGGATTTAGACTTATTATCTTAAAGAGAGAGACTGTGTGCTGC 274
547 aLeuLysGluGlyCysPheTyrValAspHisSerGlyAlaIleArgA 564
273 CCAAAAGAGAAATGTTGTTTATGTTGACCATTCAGAGAAATCAAG 224
564 sPserMetLysLeuArgLysLeuGluArgArgArgGluArg 580
223 ATCTATGCGCAAACTTAGAAGACGCTAGATATAGCTAAAGAGAAAGA 174
581 GluAlaAspGlnGlyTrpPheGlyTrpPheAsnArgSerProTrpMe 597
173 GAAAGCCAAAGAGATGTTCCGAAAGCTGTTAAATATATATATATAT 124
597 tThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuL 614
123 CACCACCTCTCTCTCCACCATAGACAGACCTTGATTACACTATGCTTT 74
614 eLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArg 630
73 TGCTTACTTTTGCCCTGATCCCTTAATTAAGTTAGTACTTTTATTTAGA 24
631 GluArgValSer 634
23 GAAAGCATTAAT 12
seq_name: gb_gss: A2430249
seq_documentation_block:
LOCUS A2430249 591 bp DNA GSS 03-OCT-2000
DEFINITION IM0214G11R Mouse 10kb plasmid U08C1M library Mus musculus genomic
clone U08C1M0214G11 R, DNA sequence.
ACCESSION A2430249
VERSION A2430249.1 GI:10554262
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0214 row: G column: 11
Seq primer: CACACGAGAAACGCTATGACC
Class: plasmid ends
High quality sequence stop: 591.
FEATURES
location/Qualifiers
1..591
source
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```

```

/clone="U08C1M0214G11"
/clone.lib="mouse 10kb plasmid U08C1M library"
/sex="Male"
/lab_host="E. coli strain XL10-gold, pi-resistant, F-"
/notes="Vector: pMD22nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD22 (g1147321149b/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-gold (Stratagene) cells
and selected for ampicillin resistance."

```

```

BASE COUNT 165 a 125 c 115 g 186 t
ORIGIN
alignment_scores:
Quality: 729.00 Length: 194
Ratio: 4.096 Gaps: 1
Percent Similarity: 91.753 Percent Identity: 71.134

```

alignment_block:

US-09-171-553B-6 x A2430249/rev ..

Align seg 1/1 to reverse of: A2430249 from: 1 to: 591

```

455 ArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLe 471
588 CCGTGGAAAGAGAACCCGTTACCTTATGAGTTTATTAAGATT 539
471 uGlyThrAlaValGlyValGlyThrAlaAlaLeuIleThrGlyP 488
538 GGGAGTACACGCTGAGTAGACAGAACCGCTGCTTATTAACACCC 489
488 tGlnGlnLeuGluLysGlyLeuGluHisAlaIleMetThrGlu 504
488 CCCAATTAATGAA.....GAGCTACGTCGACGATATGATGATT 451
505 AspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuTh 521
450 GATCTTAGACCATAGAACAGCTATATACCAATTAGAAGATCTTTAAC 401
521 rSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuP 538
400 TTCCCTATCCGAGAGTGTGCTGCAAAATAGAGGGGGTGTAGCTTATAT 351
538 heLeuArgGluGlyGlyLeuCysAlaIleLeuLysGluGlyCysPhe 554
350 TCCTTAAGAGAGGAGACTGTGCTCCCTTAAGAAAGAAATGTTGTTT 301
555 TyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArg 571
300 TATGTTGACCATTCAGAGATATCAAGATCTTATGCGTAAACCTAAGA 251
571 sLysLeuGluArgArgArgArgGluAlaAspGlnGlyTrpPheG 588
250 ACGCTAGATATATAGCTAAAGAGAAAGAGAAAGAGAGATGTTTG 201
588 uGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeu 604
200 AAAGCTGTTTATTAATAGTCCCTTGCTCACCACCTCTCTCCACTATA 151

```

```
605  TlnGlyProLeuValValLeuLeuLeuLeuTlrValGIyProCysLe 621  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
150 GCAGAGCCCTTAATTACACTTATGCTTTTGCTTACTTTGGCCCATGCAT 101  
  
621 uileasnarCPheValAlaPheValArgLnuArgValSeraIalValGlnI 638  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
100 CCTTATATAGTAGTAGAGCTTTTATTAGGAAGAAGATATAACCGACGTCAAG 51  
  
638 lmeVValLeuArgLngInlTyrgLnGLeu 648  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
50 TTATGTACTAAGCACACATATCGGTCCTT 19
```

seq_name: gb_gss:A237993

seq_documentation_block:					
LOCUS	A2337993	673 bp	DNA	GSS	29-SEP-2006
DEFINITION	cl00066FF2R Mouse 10kb plasmid UUGCIM library Mus musculus genomic				
ACCESSION	A2337993				
VERSION	A2337993.1	GI:10410826			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				

REFERENCE
AUTHORS
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
1 (bases 1 to 673)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

TITLE Islam, H., Longacre, S., Mahmud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D. Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL
COMMENT
Plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss

University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG,
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0068 row: F column: 24
Seq primer: CACACAGGAACACGCTATATACC
Class: plasmid ends
High quality sequence stop: 673.

```

FEATURES
Source
Location/Qualifiers
1. 673
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U08C1M0068F24"
/clone_lib="Mouse 10kb plasmid U08C1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/nanres/). The DNA was hydrodynamically sheared by repeated passage through 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapped DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivatized of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapped mouse DNA was annealed to adapped vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells."

```

BASE COUNT	and selected for ampicillin resistance."
197 a	120 c
167 g	188 t
1 others	

alignment_scores:	
Quality:	727.50
Ratio:	3.638
Percent Similarity:	87.719
	Length:
	Gaps:
	3
Percent Identity:	59.211

```
alignment_block:
US-09-171-553B-6 x AZ337993 . .
```

Align seg 1/1 to: AZ337993 from: 1 to: 673

```

420 ValPheasnInSerLysAspPheCysValMetValGlnIleValProAr 433
      |||||..... |||||..... |||||..... |||||.....
1  GTTTTAATTCATCATCATGATTTTGTGTCTATGATCCAGCTGTTACCCG 50
436 gValTyrTrpHisProGluGluValValLeuAsnPoluTyrAspTyrArT 455
      |||||..... |||||..... |||||..... |||||.....
51  CGTATATTATACCCCTGCATCCAGTTTA..... GAAGAAGCATATCTCG 94
453 yrasArpProLysArgGluProValSerLeuThrIleAlaValMetLeu 465
      |||||..... |||||..... |||||..... |||||.....
95  GCCGGCGCTCAAAAGAACCAATTCATTAAACCTGGCTGCATTCATG 144
470 gLyeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleTh 486
      |||||..... |||||..... |||||..... |||||.....
145  GGAAATAGTAGTGGCAGTAGAGAGTGGGGACGGAGATGCATCGTTGATAGA 196
486 rGlyProGlnGlnLeuGlnIuLysGlyLeuGlnGlyLeuHisAlaMet 503
      |||||..... |||||..... |||||..... |||||.....
195  AGGAGAGAC..... GGAAATTCAGTCTTTGAGGAGATGCTGTCA 232
503 hrGluAspLeuArgAlaLeuLysGlnSerValSerAsnLeuGlnGluSer 519
      |||||..... |||||..... |||||..... |||||.....
223  ATGAAGACCTTGACGACATAGAGAAGTCATTCAGCTTTAGAAAAATCT 282
520 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLe 366
      |||||..... |||||..... |||||..... |||||.....
283  TTGACCTCCCTGTGTGAGTGTGTTTCACAGACGAGAGAGTCTTGATTT 332
536 uLeuPheLeuArgGlnGlyGlyLeuCysAlaIleLeuLysGlnGluCysC 555
      |||||..... |||||..... |||||..... |||||.....
333  GTTGTCTTAAGAGAGAGAGAGCTGTGTCTGCTCCCTTAAAGAAAGAGTCT 382
553 yspPheTyrValAspHisSerGlyAlaIleLeuArgAspSerMetAsnLysLeu 569
      |||||..... |||||..... |||||..... |||||.....
383  GCTTCATATGACGATCAATACAGGAATAGTTAGAGACTTATGACGAACCTG 432
570 ArgLysLysLeuGlnuAArgArgArgArgGlnuArgGlnuLaspGlnGlyTr 586
      |||||..... |||||..... |||||..... |||||.....
433  AGAATAAGATTAGAGCAAGAGAAACGGGAACGGGATCTCAACGGGGGTG 482
586 pPheGlnGlyTrpPheAsnArgSerPro... TrpMetThrThrLeuLeuAs 602
      |||||..... |||||..... |||||..... |||||.....
483  GTTTGATCGTGGTTGAATCACAACACATCTGTGATTAACCTCTTTAATTT 532
602 eRalAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrValGly 618
      |||||..... |||||..... |||||..... |||||.....
533  CCGCTGTAGCGCGGACCAATCCTTATGATATGCTTAACTAGTTAGTTTCGGC 582
619 ProCysLeuIleAsnArgPheValAlaPheValArgGlnuArgLysAlaSerAl 635
      |||||..... |||||..... |||||..... |||||.....
583  CCTTGTTAATTAATTAAGCAATGGCTTTATCATCAGAGTAAATTAATGATAC 632
635 aValGlnIleMetValLeuArgGlnGlnIuTyrGln 646
      |||||..... |||||..... |||||..... |||||.....
633  AGTAAACATCATGTTCTTCAAGGCAATATCA 666

```

seq_documentation_block:	624 bp	DNA	GSS	14-DEC-2000
LOCUS	AZ642336			
DEFINITION	IM0505M18P Mouse 10kb plasmid U08C1M library	Mus musculus genomic		
ACCESSION	clone U08C1M0505M18 F, DNA sequence.			
VERSION	AZ642336			
KEYWORDS				
SOURCE	AZ642336.1	GI:11768843		
ORGANISM	GSS.			
	house mouse.			
	Mus musculus			
	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 624)			
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,			
	Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly			
	,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.			
	and Wright,D., Weiss,R.			
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb			
	plasmid inserts			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Robert B. Weiss			
	University of Utah Genome Center			
	Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT			
	84112, USA			
	Tel.: 801 585 5606			
	Fax: 801 585 7177			
	Email: ddunn@genetics.utah.edu			
	Insert Length: 10000	Std Error: 0.00		
	Plate: 0505	row: M	column: 18	
	Seq primer: CGTTGTAAGACGACGCACAT			
	Class: plasmid ends			
	High quality sequence stop: 624.			
FEATURES	Location/Qualifiers			

BASE COUNT
ORIGIN

203 a 116 c 131 g 174 t

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCLM0505M18"
/clone_1lb="Mouse 10kb plasmid UGCLM library"
/sex="Male"
/lab host="E. Coli strain XL10-Gold, Ti-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g14732114[gb]/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

alignment_scores:		
Quality:	721.50	Length: 197
Ratio:	3.964	Gaps: 2
Percent Similarity:	92.386	Percent Identity: 70.051

```
alignment_block:
```

Align seg 1/1 to: AZ642336 from: 1 to: 624

```

460  pvalaSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValG 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3  CCCGTTACCTTACTTTGGCAGTTCATTAGAGATTGGAGACGACTGG 52
476  yValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuL 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53  ACTAGGTACAGGACCGCTGCCTTAATTATAGACCCCAACTACTATGAA 101
493  yGcGlyLeuGlyGlnLeuHisAlaAlaMetThrGluAspLeuAlaLeu 509
|||||:|||||:|||||:|||||:|||||:|||||:
102  ..... GAACCTAGCTGACGATGATGATTGATCTTAGAACATA 140
510  LysGlnSerValSerAsnLeuGlnGlnSerLeuThrSerLeuSergLu 526
|||||:|||||:|||||:|||||:|||||:|||||:
141  GACACGCTCTATACCAATTTAGAGAAATCTTTAACTTCCCTGCCGAGT 190
526  lValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGly 543
|||||:|||||:|||||:|||||:|||||:|||||:
191  GGTGCTGCAGAAATAGAAAGGGATTAGACTTATTATTCCTTAAGAAAG 240
543  lYLeuGlyAlaAlaLeuGlyGlnGlyCysGlyPheTyValAspHisSer 559
|||||:|||||:|||||:|||||:|||||:|||||:
241  GACCTGTGCTGCCTTAAAGAACAGATGTTGTTTATGTTGACCAATCA 290
560  GAlaAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuAla 576
|||||:|||||:|||||:|||||:|||||:|||||:
291  GGAGTAATCAAGATCTATGGCTTAACTTGAAGAACGCCATGATATAGC 340
576  gArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPheAsn 593
|||||:|||||:|||||:|||||:|||||:|||||:
341  TAAAGACAGAAAGAAACCAACAGATGTTTGAAGACGCGTTTATA 390
593  rGSerProTyrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuVal 609
|||||:|||||:|||||:|||||:|||||:|||||:
391  AGTCCCTTGGCTCACACACTCTCCTCCACTATAGACGACCTTAAAT 440
610  ValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheVa 626
|||||:|||||:|||||:|||||:|||||:|||||:
441  ACACCTTAAGCTTTGGCTTACTTTGGCCCAAGCATCCCTATATAGTTAGT 490
626  lAlaPheValArgGluArgValSerAlaValAlaGlnIleMetValLeuArg 643
|||||:|||||:|||||:|||||:|||||:|||||:
491  AGCTTTTATTAGACAAAGATAAACCGACGTCCAAAGTTATGCTCCTTAGAC 540
643  lnglntyrGlnGlyLeuLeuSerGlnGlnGlyIleThrAspLeu 656
|||||:|||||:|||||:|||||:|||||:|||||:
541  AACATATTCGG...GTCTTCAGAGAGGTGAAACACTCGCTC 578
seq_name: gb_gss:A2741412

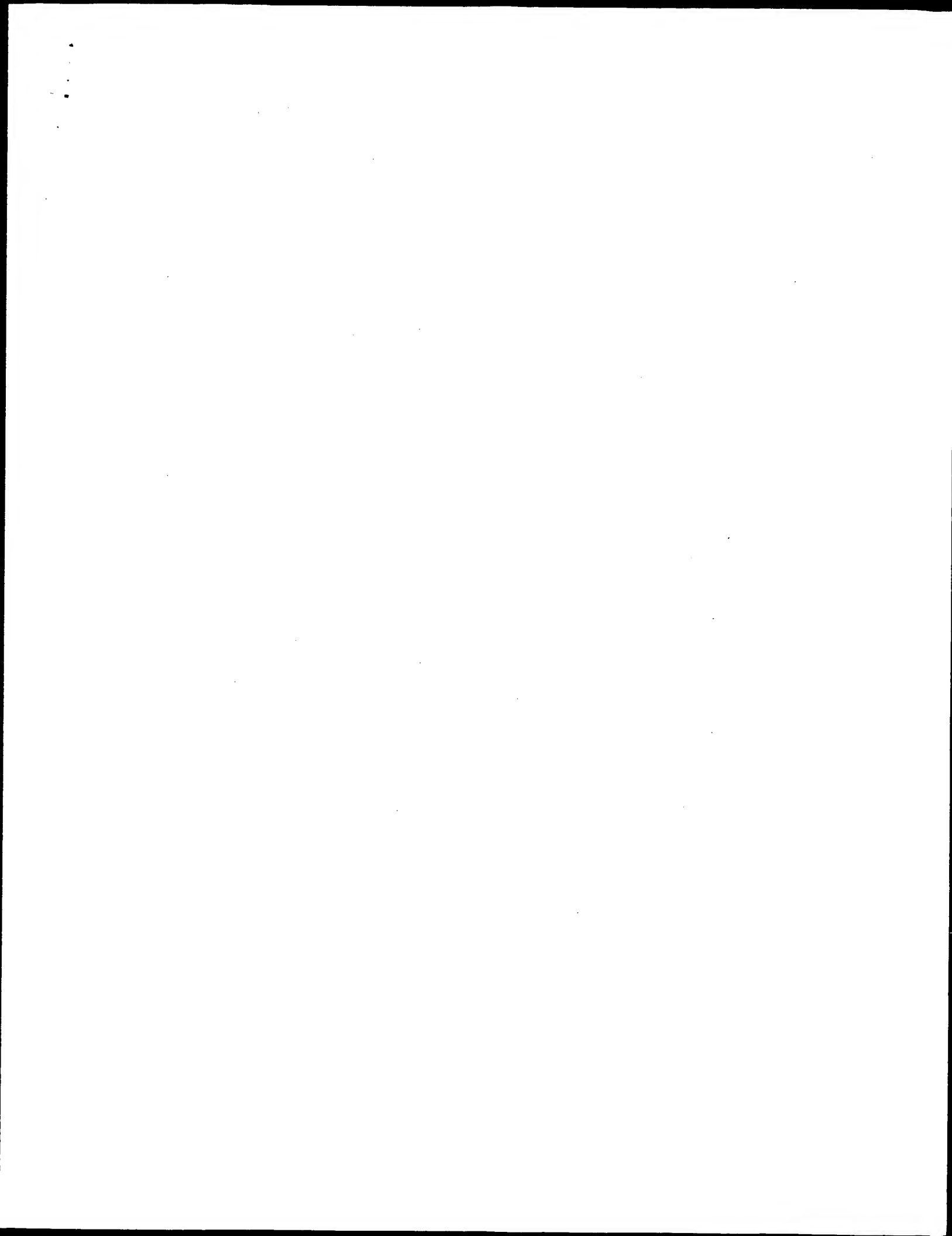
```

seq_documentation_block:	659 bp	DNA	GSS	25-JAN-2001
LOCUS	AZ741412			
DEFINITION	RPC1-24-74C13_TV RPC1-24	Mus musculus	genomic clone	RPC1-24-74C13
ACCESSION	AZ741412			
VERSION	AZ741412.1	GI:12518156		
KEYWORDS	GSS.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus 1 (bases 1 to 659).			
AUTHORS	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akhurst,B., Levins,M., Tsegaye,G., Geer,K., Kroll,M., Shwartsbeyn,A., Gebregorgis,E., Russell,D., de Jong,P. and Fraser,C.N.			
TITLE	Mouse BAC End Sequences from Library RPC1-24			
JOURNAL	Unpublished (1999)			
COMMENT	Other_GSSs: RPC1-24-74C13_TV			

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Mon Feb 25 07:44:30 2002

us-09-171-553b-6.feb22std.rst



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:03 ; Search time 6550.43 Seconds
(without alignments)
4849.227 Million cell updates/sec

Title: US-09-171-553B-9
Perfect score: 2956
Sequence: 1 tgccttttaggttagaac.....aaaaaaaaaaaaaa 2956

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 537289281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rpd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472	16.0	549	10	AM657531 110922 MA
2	383.6	13.0	564	11	BI182930 UNC-P-FN-
3	358.6	12.1	602	11	BF712162 MT-P-E6-a
4	293.6	9.9	754	13	A2970955 2M0244H16
5	271.2	9.2	494	11	BI182742 UNC-P-FN-
6	258.6	8.7	701	13	A2351242 1M0089L09
7	255	8.6	774	13	BI122287 RPT-24-2
8	254.4	8.6	591	13	A2430249 1M0214G11
9	253	8.6	632	13	A2814037 2M0081N06
10	246.8	8.3	624	13	A2642336 1M0505M18
11	245.2	8.3	641	13	A2850764 2M0152C18
12	242.6	8.2	493	13	A2349077 1M0085K21

Result No.	Score	Query Match	Length	ID	Description
13	240.8	8.1	650	13	A2620796 1M0453112
14	240.6	8.1	625	13	A2813042 2M0080E11
15	234.6	7.9	592	13	A2836232 2M0131E01
16	232	7.8	534	13	A2499035 1M0336P05
17	230.6	7.8	610	13	A2376517 1M0336P05
18	229.4	7.8	644	13	A2356870 1M0098F09
19	228.6	7.7	609	13	A2439189 1M0229N10
20	228	7.7	730	11	BI185535 UNC-P-FN-
21	227.6	7.7	656	13	A2375836 1M0129J20
22	225	7.6	675	13	A2739547 RPT-24-7
23	224	7.6	600	13	A2987182 2M0269J09
24	222.2	7.5	895	10	BE569478 601328494
25	221.6	7.5	597	13	A2406484 1M0175C17
26	221.6	7.5	882	13	BE703805 MT-P-O2-a
27	220	7.4	576	13	A2730616 RPT-24-3
28	219	7.4	834	13	BI124666 RPT-24-2
29	215.4	7.3	583	13	A2417785 1M0193B08
30	214.2	7.2	669	13	BE703805 MT-P-O2-a
31	214	7.2	554	11	BF703805 MT-P-O2-a
32	211.2	7.1	659	13	A2741412 RPT-24-7
33	211	7.1	485	11	BF703671 MT-P-E4-a
34	210.6	7.1	704	13	BI120896 RPT-24-2
35	209.6	7.1	461	10	AL120896 DFE2P62N
36	209	7.1	640	13	BE703805 MT-P-O2-a
37	209	7.1	679	13	A2647666 1M0514J12
38	208.8	7.1	636	13	A2727116 RPT-24-1
39	208	7.0	628	13	A2406003 1M0175N06
40	207.2	7.0	906	11	BI328061 602979695
41	207	7.0	668	13	A2327473 1M0050K14
42	200.6	6.8	631	11	BF702741 MT-P-E3-a
43	198.8	6.7	442	13	A2597961 1M0412K22
44	198.2	6.7	625	13	A2336017 1M0066P05
45	197.8	6.7	605	13	A2416306 1M0191D01

ALIGNMENTS

RESULT 1
LOCUS AM657531 549 bp mRNA EST 25-APR-2001
DEFINITION 110922 MARC 1BOV Bos laurus cDNA 5', mRNA sequence.
ACCESSION AM657531
VERSION AM657531.1 GI:7423429
KEYWORDS EST.
SOURCE cDNA
ORGANISM Bos laurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 549)
AUTHORS Smith,T.P.L., Grosse,W.M., Fekling,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pertea,G., Holt,I., Karanymcheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smitht@mail.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCAGCAGC
Plate: 94 row: E column: 12

Db 264 GGGCTCTCGATATTTTAAATGATTTGGTCCACGAGCGGGCTCTCGATATTTTAAAT 205
Qy 2768 -----"gactcgatatttaaatgattgattgattgacacag 2808
|||||
Db 204 GATTGTCACGAGCGCGCTCTCGATATTTTAAATGATTTGGTGGACGACAGC 145
Qy 2809 cttgtgtgaacccataaagctgtccgactcgcacgtcgagcgagcagctctac 2868
|||||
Db 144 CTTGTGTGTAACCCATAAAGCTGTGCGACTCCGACACAGCGCGCGAGTCTCTAC 85
Qy 2869 cccgtgtgtgtacgactgtggcccgagcgagcgttggataaactctctgtct 2928
|||||
Db 84 CCGTGTGGCTACGACTGTGGCGCCGACGCGCTTGAATTAATCTTGTGCTT 25
Qy 2929 tgcatacaaaaaaaaaaaaaa 2952
|||||
Db 24 TGCATCAAAAAAAAAAAAAA 1

RESULT 3
BF712162/c 602 bp mRNA EST 02-JAN-2001
LOCUS BF712162
DEFINITION MI-P-E6-acx-e-11-1-UM.s1 MI-P-E6 Sus scrofa cDNA clone
MI-P-E6-acx-e-11-1-UM 3', mRNA sequence.
ACCESSION BF712162
VERSION BF712162.1 GI:12011595
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 602)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Tugale CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kilde Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: ktugale@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized fetus at gestational day 20 library cDNA library
Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
Research Center, Department of Animal Science, University of
Missouri-Columbia, 65211 Clone distribution: clones will be
available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source location/Qualifiers

1..602
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-E6-acx-e-11-1-UM"
/clone_lib="MI-P-E6"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E6
library is derived from fetus at gestational day 20. For
a detailed description of the library from which this
clone was derived, please visit our web site at
http://pigest.genome.iastate.edu/.
TAG_LIB=MI-P-E6
TAG_TISSUE=fetus at gestational day 20
TAG_SEQ=AGGAA"

BASE COUNT 141 a 146 c 152 g 162 t 1 others
ORIGIN

Query Match 12.1%; Score 358.6; DB 11; Length 602;
Best Local Similarity 72.9%; Pred. No. 4,4e-67;
Matches 557; Conservative 0; Mismatches 45; Indels 162; Gaps 1;

Qy 2186 cccatagtagctgcgcctccgcttactaagttgggctgtgcttaataatagttt 2245
|||||
Db 602 CCTTAGTGGCTCTGCTCTGCTGCTTACAGTTGGCGCTTCTTAATTAATAGGTTGGT 543
Qy 2246 cctttagaagaacagagtgagtgagtcagtcacatcatgttacttaggaacag 2305
|||||
Db 542 CCNTGTTAGAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 483
Qy 2306 gctcttgagccaaggaagaactgacctagcctccagttcagtttaagataag 2365
|||||
Db 482 GCCTTCGAGCCAG 423
Qy 2366 aacacagaagaagtgaggaggaatggaatggaatggaatggaatggaatggaatgga 2425
|||||
Db 422 CACAGACAG 363
Qy 2426 ggaagtaataaagaagcttaaatgccccgaattacagaacctgtgctgcagtaa 2485
|||||
Db 362 AGAGGTAGTGAAG 332
Qy 2486 taggtagaagtgacacatctcctatgttccagggcctgctatcctgagtaagata 2545
|||||
Db 333 ----- 332
Qy 2546 aacagaatgagtgactaactatgcttctgattctgtaaaactgacgacacataga 2605
|||||
Db 333 ----- 332
Qy 2606 agaattgatacactgacagccctagtgactatcactgaactgaactgactctgc 2665
|||||
Db 333 -----TTCCATAAAG 285
Qy 2666 ccaagagcccaagcagagtgagagctcggagctatcttaaatgattggtccagagc 2725
|||||
Db 284 CCAG 225
Qy 2726 gggggctcgaattatttaaatgattggtcgaatgagagcgggctcgaatttaa 2785
|||||
Db 224 GGGGCTCTCGATTTTAAATGATTTGGTCCAGGAGACGGGCTCTCGATTTTAA 165
Qy 2786 atgattgtttagtcagcagagctgtgtgtaacccataaagaactgtccagatccg 2845
|||||
Db 164 ATGATTGTTGGTTCGAAG 105
Qy 2846 caatcgaggcagcagatcctcctacccctgcgtgtgttgaagctgtgagcccaagcgct 2905
|||||
Db 104 CACTCGGGCGCCAG 45
Qy 2906 ggaataaataatcctctgtgttgcattcaaaaaaataaataaataaataaataa 2949
|||||
Db 44 GGATTAATAATCTCTGCTGTTGATCAAAAAAAAAAAAAA 1

RESULT 4
A2970955/c 754 bp DNA GSS 27-APR-2001
LOCUS A2970955
DEFINITION CMO244H16F Mouse 10kb plasmid U06C2M library Mus musculus genomic
clone U06C2M0244H16 F, DNA sequence.
ACCESSION A2970955
VERSION A2970955.1 GI:13842182
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 754)
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamll, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Kelly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A., and Wright, D., Weiss, R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., STC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddomegenetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0244 row: H column: 16
 Seq primer: CGTGTAAACGACGCGCCACT
 Class: plasmid ends
 High quality sequence stop: 754.

FEATURES

Location/Qualifiers
 1..754
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUC2M0244H16"
 /clone_id="Mouse 10kb plasmid UUC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv: Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114pb1AP129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 226 a 160 c 150 g 218 t
 ORIGIN

Query Match 9.9%; Score 293.6; DB 13; Length 754;
 Best Local Similarity 64.8%; Pred. No. 4.2e-53;
 Matches 474; Conservative 0; Mismatches 244; Indels 14; Gaps 2;

QY 1882 tggagcagcaactggtggttaacccctggttccacccagtcctcaacaaatccaa 1641
 Db 751 TGGGCTTGCAATACCGGTCATCTCTGTCTATGCTATTTAATAGTCCAA 692
 QY 1642 gattgtgtcatggtcccaatgctcccgagtgctactcatctctggaagtgctc 1701
 Db 691 GATTTCGCAATTTGGTTCAGCTTATTCCTAGACTCTGTATCATGATGATAGCTCTCTT 632
 QY 1702 ctgtatgaatgactctggtataacgcaaaagaaagaaacccgtatcccttaccta 1761
 Db 631 TTAGACAAATTTAGCATCG---GTCCGCTGAGAGAGAGAGAGCCCATTTACTTTG 575
 QY 1762 gctgtaatgctggtatgaagcgcgtgtgctggaacaggaagcagctgcctgac 1821
 Db 574 ACAAGTCTATTAGGATTAGAGTAGCGGTGAGATAGGTACAGAGACCGCTGCTTAAT 515

QY 1822 aacagaccacagcagctagagaaggaactgtgtgactacatcgccatgacagaagt 1881
 Db 514 AAGACCCCAAAATCTAA-----TGAGAACTACGTCCACTATGATGTTGAT 466
 QY 1882 ctccagccttaagagagctgttgaacactgaagagagcctcctgctctgtgtga 1941
 Db 465 CTTAGAAGCTATGAGACAGTATATACCAATTTGAGAAATCTTAACTTCCCTCGAA 406
 QY 1942 gtgttctacagaaccggaaggagattagatctgtcttcttaagaagaagtgtgt 2001
 Db 405 GTGAGTCTACAGATAGAAAGGAGATTAGACTTATTTATTCCTTAAGAGAGAGACTCTG 346
 QY 2002 gcaagccttaagaagaatgtgtcttcttctatgtagatcactcaggaacatgaagctcc 2061
 Db 345 GCTGCCCTTAAAGAGATGTTGTTTATGTTGACCATTCAGCAATATCAAAAGTTCT 286
 QY 2062 atgaacaagcttgaagaagaagtagagaagcgtcgaagggaagaagagcgtacccaggg 2121
 Db 285 ATGGCCAACTTATGAGACAGCCCTAGATACAGCTAAAGGAAAGAAAGCAACAGGA 226
 QY 2122 tggcttgaagatgtgtcaacaggtctccttgaatgacacccctgtctgtcagc 2181
 Db 225 TGCTTGAAAGCTGTGTTAATAGTCCCTTGCTCACCACCTCTCCACCATAGCA 166
 QY 2182 gggccctgtgtgtctgtcctcctgtactacagttggtgctgtcttaattagttt 2241
 Db 165 GGACCTTGTGATCTCTTATAGCTTTGCTTACTTTTGCCCTTCGATCTCTTAATAGTTA 106
 QY 2242 gtgccttgttagagaagagtgagtgagtcagtcagcatcatgttacttaggaacagttc 2301
 Db 105 GTAGCTTTTATTTAGAAAGAGATATGACAGTACAGATTATGTACTAAGCAACATAT 46
 QY 2302 caagcctctgt 2313
 Db 45 CGGCTCTTACAG 34

RESULT 5
 B1182742/c
 LOCUS
 DEFINITION
 UNL-P-FN-do-e-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 UNL-P-FN-do-e-05-0-UNL 3', mRNA sequence.
 ACCESSION B1182742
 B1182742.1 GI:14657151
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 494)
 Caetano, A.R., Johnson, R.K. and Pomp, D.
 Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 Unpublished (2001)
 CONTACT: Pomp, D
 DEPARTMENT: Department of Animal Science
 UNIVERSITY: University of Nebraska, Lincoln
 LINCOLN, NE 68583-0908, USA
 TEL: 402 472 6416
 FAX: 402 472 6362
 EMAIL: ddomp@unl.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 normalized porcine ovarian follicles library
 Seq primer: M13 -29
 POLYA=Yes.

FEATURES

Location/Qualifiers
 1..494
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection

RESULT	8
AZ430249/c	
LOCUS	AZ430249 591 bp DNA GSS 03-OCT-2000
DEFINITION	M0214G1R Mouse 10kb plasmid UUCCLM library Mus musculus genomic clone UUCCLM0214g1 R, DNA sequence.
ACCESSION	AZ430249
VERSION	AZ430249.1 GI:10554262
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 591)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmood,M., Meenen,E., Petersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid insets
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0214 Row: G Column: 11
Seq primer: CACACAGGAAACACGCTAGACC
Class: plasmid ends
High quality sequence stop: 591.
Location/Qualifiers

BASE COUNT
ORIGIN

165 a 125 c 115 g 166 t

```
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db_xref="taxon:10090"  
/clone="U08C1M0214G11"  
/clone_1fb="Mouse 10kb plasmid U08C1M library"  
/sex="Male"  
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PMD42 (g11473211419b|AF192072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."
```

	8.6%;	Score 254.4;	DB 13;	Length 591;	
Query Match	Best Local Similarity	66.2%;	Pred. No.	1.2e-44;	
Matches	388;	Conservative	0;	Mismatches	186;
				Indels	12;
				Gaps	1
Cy	1728	cgcacaaagaagaccgctatcccttacctagtctgaatcgttcgtaattggcgagcgc	1787		
Dd	589	ccccggaanaadgaaccccgtttaccttacttttgccaggtttatTTTnDGGATTGGAAGTAAC	530		

QY	1788	cgttggcgttaagaaacagagacagctcggccctatcacagagacacagcgcgttagaagaaag	1847
Db	529	AGCTGGAGTACGTACGAGAACCGCGCTTAAATTAACACCCGCCAAATCTATGAA-----	473
QY	1848	acttgatgagctacatgcgcgcacatgacagaagaatccgagacctaaaggagctcgttag	1907
Db	474	-----GAGCTACGAGTCGAGCTATGAGANGATTGATCTTTGAAACCAATAGAACATCTATATAC	422
QY	1908	caactagaagaagtcoccgacctctcttgcgaagtgtttcaagaacctgaagaaccgaggagat	1967
Db	421	CAAAATTAGAAATCTTTAACTTCCCTATCCGAAATGCTGCTGCAGAAATTAGAGGGGGTT	362
QY	1968	agatcgcgtcttctcaagagaagtggttlatgtgcagccctlaaagaagaatlgctt	2027
Db	361	AGACTTATTTATTCCTTAAAGAGGAGGAGCATCGTCTGCTCCCTTAAAGAAAGAAATGTTGATT	302
QY	2028	ctatgtatatactacagagacatcagagacatccatgaagacccatgaacaaagcttagaaaaagtga	2087
Db	301	TTATGTTTACCACTTCAGGAGTATATAAGATTTCTATGGCTAACTTTAANGAACGCTTAA	242
QY	2088	gagcgctcgaagagaaagagagagctctgacacaaaggtggtttgaagatagtltcaacagtlc	2147
Db	241	TATAGCTAAAGAGAGAAAGAGAAAGCCCAAGAGATGGTTTGAAACCTGGTTTAAATAGTC	182
QY	2148	tccttgaatgacacaccctgcttctgcgtctgaaggggccccctagtagcctgctcgtt	2207
Db	181	CCCTTGGCTACCACTCTCTCTCCACATATAGCAAGACCTTTAATTACACTATGCTTTT	122
QY	2208	acttaacgtltgggccttgcttaatctaaatlaagttgttgcctlttgttgaagaacgagtag	2267
Db	121	GCTTACTTTTGGCCCATGATCCTTAATAAGTTAGTAGCTTTTATTAGAGAAAGATATA	62
QY	2268	tgcagtcagatatactgttaacttagagcaacagatcaacaaggcctctg	2313
Db	61	CGCACTCCCAAGTTATGTGACTAAGGCAACAATATAGGGCTCTTACG	16

```

RESULT      9
LOCUS       AZ814037
DEFINITION  AZ814037 632 bp DNA GSS 20-FEB-2001
            2M0081IN06R Mouse 10kb plasmid UGCC1M library Mus musculus genomic
            clone UUGC2M0081IN06 R, DNA sequence.
ACCESSION   AZ814037
VERSION     AZ814037.1 GI:12983945
KEYWORDS
SOURCE      house mouse.
            Mus musculus
ORGANISM    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 632)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
JOURNAL
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLc, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: dduwn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0081 row: N column: 06
            Seq primer: CACACAGCAACACGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 632.
            Location/Qualifiers
                1..632
                /organism="Mus musculus"

```


QY 1802 caaggagacgtccctgcatcacagagaccagcagctagagaagactggtgagctac 1861
 DB 61 CAGGAACCCGCTGCTTAAATAGACCCCAATAGTAAGAA-----CTAC 108
 QY 1862 atggagccatgacagaagactccagccttaagaaggtctgttagcaactagaagact 1921
 DB 109 GTGACGCTATGATATGATCTTATGACTATGACACGCTTATACCAATTTGAGGAAT 168
 QY 1922 cccgactcttctgctgaagtggtctacagaaccgagggagattagatctgctgttc 1981
 DB 169 CTTTAATCTTCCGCTGCGAAGTGTGCTGCAAAATAGAGGGGATGACTATATATTC 228
 QY 1982 taagaagaagtgggtatgtgacagccttaagaagaagtgtctttagttagtact 2041
 DB 229 TTAAGAAGAGGAGGACTGCTGCTGCTTAAAGAGATGTGTTTATGATGACCAT 288
 QY 2042 caggagccatcagagactccatgacaagcttaagaagaagttagagaagcgtcgaag 2101
 DB 289 CAGAGATATTAAGATTCTATGCTAACTTAGAGAACCCCTAGATATACGTAAGAG 348
 QY 2102 aaagaagcgtcagaccaggggtgttgaagatgttcaacaggtctccttgatgaca 2161
 DB 349 AAAGGAAAGCCCAACAGAGATGTTGAAGCTGTTTAAATAGTCCCTTGCTGCTACA 408
 QY 2162 cccgtcttctgctcagcaggggcccctagtagtccctgctcctgttactacagtggc 2221
 DB 409 CTCCTCTCCATATAGCAGACCTTAAATACCTTATGCTTCTTCTTCTTCTTGGCC 468
 QY 2222 ctgtcttaataataggttcttcccttctgttagagaacaggtgagcagctccagatca 2281
 DB 469 CATGATCTTATATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 528
 QY 2282 tggacttaggcaacagctacccaagcctctctg 2313
 DB 529 TGGTCTTAAGACAAATATGCGGTCTTCAAG 560

RESULT 11
 A2850764/c
 LOCUS 2M0152C18R Mouse 10kb plasmid UNGC1M library Mus musculus genomic
 DEFINITION clone UNGC2M0152C18 R, DNA sequence.
 ACCESSION A2850764
 VERSION A2850764.1 GI:13036088
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 641)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0152 row: C column: 18
 Seq primer: CACACAGAAACAGCTATGACG
 Class: plasmid ends
 High quality sequence stop: 641.
 Location/Qualifiers
 1..641
 source

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UNG2M0152C18"
 /clone_11b="Mouse 10kb plasmid UNGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD24hv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 ligated DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g147321419b1AF129072.1), a copy number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT

197 a 134 c 123 g 187 t

ORIGIN

Query Match 8.3%; Score 245.2; DB 13; Length 641;
 Best Local Similarity 63.5%; Pred. No. 1,1e-42;
 Matches 414; Conservative 0; Mismatches 223; Indels 15; Gaps 2;

QY 1610 ggtgtccaccctagcttccatcccaatgaattgtgtgcataggtcccaatgctcc 1669
 DB 641 GTGTGCTATGCTGTGTTTAAATAGTCCAAATTTGTCATTTGCTTACCTTATTC 582
 QY 1670 cccgagtgatctaccctccctcaggaagtggtcttgatgatagtactacgtataacc 1729
 DB 581 CTAGACTCCCTGATATCATGATGATGATGATGATGATGATGATGATGATGATGAT 525
 QY 1730 gaccataaagaagaccgtatccctaccctagctgtaatgctcgattagaagcggcg 1789
 DB 524 GCTGAGAAAGAAAGAACCCGTTACTTACTTGTGACAGTCTTATGAGATTGGAGTAGCGG 465
 QY 1790 ttggcgttagaacaaggaagcgtgcccctgctacagagaccagcagcgttagaagagac 1849
 DB 464 CTGAGATAGGTACAGGAACCGCTGCTTAAATTAAGACCCCTCC-----AATACT 417
 QY 1850 ttgttgatcatagatgagcagatgacagaatctccagccttaagaagatctgttagca 1909
 DB 416 ATGAGAACTACGTCGACGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 357
 QY 1910 acctagaagatccctgactcttctgtcaagtgttctacagaaccgagagggattag 1969
 DB 356 AATTAGAAGATCTTTAACTTCCCTGCGCAAGGTGTGTAACAGAAATGAAGAGGAGATTAG 297
 QY 1970 atctgctgttcttaagaagagtggtatgtgtagccttaagaagaagtgttcttct 2029
 DB 296 ACTTATATTTCTTAAAGAGAGGAGCTGTGCTGCTTAAAGAGAGAAATGTTGTTT 237
 QY 2030 atgtagatcatcagagacatcagagatccatgacaagccttagaanaaagttagaga 2089
 DB 236 ATGTTGACCATTTAGAGATATCAAAAGATTCTAGGCAAACTTAGAAGACGCTTACATA 177
 QY 2090 ggcgtcgaaaggaagagcgtgacaggggtgttgaagatggttcaacaggtctc 2149
 DB 176 TACTTAAAGAGAAAGAAAGCCACAGAGATGTTGCAAAAGCTGTTTAAATAGATTCC 117
 QY 2150 cttagagacacccctcttctgctgacagggcccctagtagtctgctcgtctgac 2209
 DB 116 CTGGCTCACCACTCTCTTCCACATAGCAGAGACCTTATATACATTATGCTTTTGC 57

```
/db_xref="taxon:10090"
```

/clone="UUCG1M0453112"
 /clone_lib="Mouse 10kb plasmid UUCG1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 180 a 131 c 128 g 211 t
 ORIGIN

Query Match 8.1%; Score 240.8; DB 13; Length 650;
 Best Local Similarity 67.0%; Pred. No. 1e-41;
 Matches 363; Conservative 0; Mismatches 167; Indels 12; Gaps 1;

QY 1772 tcgattaggaagccgctggtggtggaagcagcagctgcccgtacacagcacc 1831
 DB 649 TAGAGTTGGGAGTAGCAGCTGAGTACGAGAAATGCGTTCCTTAATTAAGACCCCC 590
 QY 1832 agcagctagagaagagactgtgtgacatagcgcagatagagaagatccagcct 1891
 DB 589 CAATCTATA-----AGAACTACGTCGACGTATGATGATGATGATGATGATGAT 542
 QY 1892 taagagagctgttagcaactagagaagatccgtactcttctgctgaagtggtttac 1951
 DB 541 TAGAAGAGTCTATACCAAAATTAAGAAATCTTAACTCCCTCCGAAGTGTGTGC 482
 QY 1992 agaaccgagagagtagatgctgttcttaagagaagtggtttagtgcacccctaa 2011
 DB 481 AAAATAGAGGAGATAGACTTATTTCTTAAGAGAGAGAGACTGTGCTGCTTAA 422
 QY 2012 aagaagaatgttctatgtatgactacactcagagcactcagactcagacgaacg 2071
 DB 421 AAGAAGATGTTGTTTATATGTCACCTCAGAGATATCAAGATTTATGCGCAAC 362
 QY 2072 ttagaagaagttgagagagcgtcgaagaggaagagagcgtgaccagagtggttaag 2131
 DB 361 TTAGGAAGAGCGCTAGATTTAGCTAAAGAGAAAGCAAGCAAGATGATTGAAA 302
 QY 2132 gatgttaacaggtctccttgatgacacccctgcttctcctcagagggccctag 2191
 DB 301 GCTGTTTAAATAGTCCCTTGCTCAGCACTCTCTCCATGATAGAGAGACTTTAA 242
 QY 2192 tagtctgctcctgttactacagttgggctgtcctaataatagttgttgccttg 2251
 DB 241 TTACACTTATGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTTAA 182
 QY 2252 ttagagaagagtgatgacatcagtgacttagcacaagagccttc 2311
 DB 181 TTAGGAAGAGATTAAGCAGTCAAGTTATGATGATGATGATGATGATGATGATGAT 122
 QY 2312 tg 2313
 DB 121 AG 120

RESULT 14

AZ813042/c
 LOCUS AZ813042 625 bp DNA GSS 20-FEB-2001
 DEFINITION 2M0080E11F Mouse 10kb plasmid UUCG1M library Mus musculus genomic clone UUCG2M0080E11 F, DNA sequence.

ACCESSION AZ813042
 VERSION AZ813042.1 GI:12982881
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS
 TITLE
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus-1 (bases 1 to 625)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL
 COMMENT
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert Length: 1000 Std Error: 0.00
 Plate: 0080 row: E column: 11
 Seq primer: CTTGTAAACGACGCCAGT
 Class: plasmid ends
 High quality sequence stop: 625.
 Location/Qualifiers

FEATURES

1..625
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCG2M0080E11"
 /clone_lib="Mouse 10kb plasmid UUCG1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 191 a 128 c 120 g 186 t
 ORIGIN

Query Match 8.1%; Score 240.6; DB 13; Length 625;
 Best Local Similarity 63.8%; Pred. No. 1.1e-41;
 Matches 404; Conservative 0; Mismatches 214; Indels 15; Gaps 2;

QY 1574 acaggtggtggatgacatcaggttaaccccggtgttccacctcagttcaacc 1633
 DB 618 ACACAGTGTGAGCTTCAATAGCGTCTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 559
 QY 1634 aatccaaagtgtgtgtatgttccaaatcgtcccccagagtgatctacacccctgagg 1693

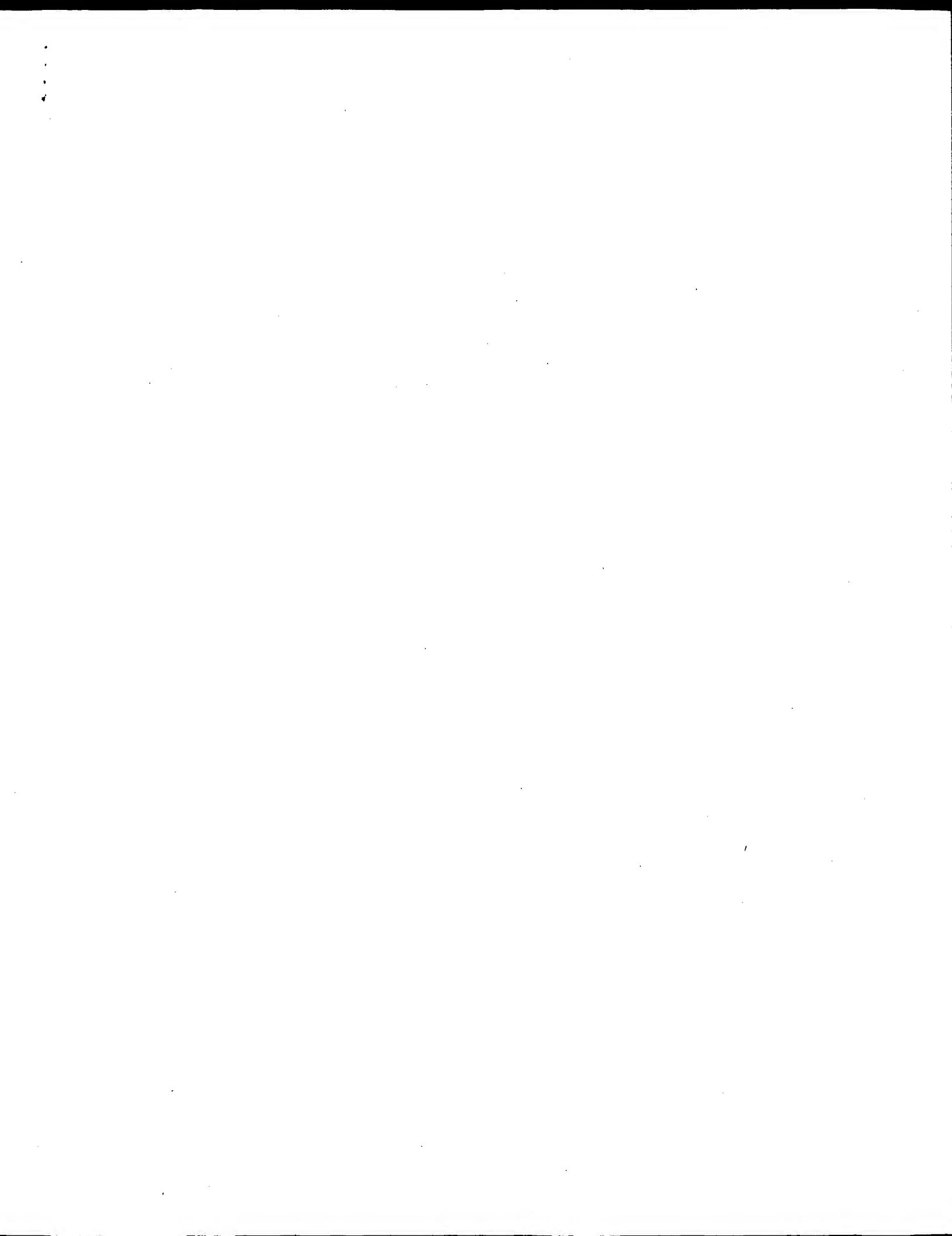
• Mon. Feb 25 07:44:34 2002

us-09-171-553b-9.feb22std.rst

Page 13

|||||||
Db 562 TGCAGTCCAAGTC 574

Search completed: February 23, 2002, 23:37:18
Job time: 9439 sec



OM of: US-09-171-553b-10 to: EST.* out_format: pfs

Date: Feb 24, 2002 8:32 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framed_p2n.model -DEV=x1h
-q/cg22.1/USPTO_spool/US09171533/runat_22022002.122825.11137/app_query.fasta.1.3312
-b=EST -QFMT=fastap -SUFFIX=fe22std.rst -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-EGAPOP=6.000 -EGAPEXT=7.000 -VGAPOP=10.000 -VGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09171533_CG22.1.9008 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-171-553b-10
Query length: 657
Database: EST.*
Database sequences: 1131937
Database length: 1077921985
Search time (sec): 3515.700000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Document
gb_est1:AM657531	+	888.00	1216.45	1.4e-58	549
gb_gss:AZ270955	-	875.50	1196.21	1.8e-57	734
gb_gss:AH059027	-	859.50	1172.93	3.7e-56	882
gb_gss:BH122827	+	843.50	1151.93	3.7e-55	774
gb_gss:AZ251442	-	814.00	1112.30	8.6e-53	701
gb_est:BI151181	+	787.00	1073.36	1.3e-50	845
gb_gss:AZ850764	-	770.50	1053.32	1.7e-49	641
gb_gss:AZ813042	-	757.50	1035.67	1.6e-48	625
gb_est:BI106669	+	757.00	1031.27	2.8e-48	921
gb_gss:AZ256870	-	734.50	1003.74	9.6e-47	644
gb_gss:AZ2376517	-	730.50	998.76	1.8e-46	610
gb_gss:AZ2430249	-	729.00	997.00	2.3e-46	591
gb_gss:AZ242336	+	721.50	986.16	9.1e-46	624
gb_gss:AZ237993	+	721.50	985.44	1.0e-45	673
gb_gss:AZ214037	+	713.50	975.04	3.8e-45	632
gb_gss:AZ214142	+	711.00	971.07	6.3e-45	659
gb_gss:AZ237473	-	696.50	952.72	6.7e-44	565
gb_gss:AZ295836	+	693.00	948.45	1.2e-43	534
gb_gss:AZ239547	-	693.00	946.19	1.5e-43	676
gb_gss:AZ217785	-	685.50	937.29	4.8e-43	583
gb_est:BI109841	+	681.50	928.91	1.4e-42	787
gb_gss:AZ2620796	-	671.50	916.99	6.5e-42	650
gb_gss:AZ2836232	+	670.00	915.82	7.6e-42	592
gb_est:BF608155	+	667.00	912.54	1.2e-41	833
gb_est:BF56478	+	666.00	910.48	1.5e-41	895
gb_gss:AZ216306	-	662.50	905.29	2.9e-41	605
gb_est:BF159914	+	659.00	899.36	6.2e-41	680
gb_gss:BH054319	-	653.50	890.50	1.9e-40	724
gb_gss:AZ236017	-	652.50	891.22	1.8e-40	625
gb_est:BF181106	+	645.00	876.70	1.1e-39	969
gb_gss:AZ2764389	+	642.50	877.72	1.0e-39	609
gb_gss:AZ2647666	+	638.00	870.48	2.5e-39	679
gb_est:BI106591	+	628.50	853.99	2.1e-38	728
gb_est:BI158931	+	624.50	852.24	2.5e-38	716
gb_gss:AZ2989131	+	624.50	852.24	2.5e-38	656
gb_gss:AZ249077	-	619.00	847.41	4.9e-38	493
gb_est:BI1082444	+	618.00	842.01	9.8e-38	750
gb_gss:AZ2960032	-	612.00	835.02	2.3e-37	644
gb_gss:AZ2397096	-	611.50	834.08	2.7e-37	675

```
gb_gss:AZ296293 - 603.50 824.77 8.9e-37 566 | AZ296293 2M0282119F Mouse
gb_est2:BF786434 - 602.00 819.37 1.8e-36 801 | BF786434 602112966F1 NCI_C
gb_gss:AZ2887402 - 599.00 818.42 2.0e-36 575 | AZ2887402 RPT-24-176N10.TJ
gb_gss:BH055839 + 596.00 813.57 3.7e-36 620 | BH055839 RPT-24-93P16.TVB
gb_gss:AZ637496 - 596.00 813.08 4.0e-36 653 | AZ637496 1M0496123R Mouse
```

seq_name: gb_est1:AM657531

seq_documentation block:

```
LOCUS AM657531 549 bp mRNA EST 25-APR-2001
DEFINITION 110922 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION AM657531
VERSION AM657531.1 GI:7423429
KEYWORDS EST.
SOURCE EST.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 549)

```

REFERENCE Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Kohrer,G.A., Chitko-Mckown,C.G., Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine CDNA libraries and construction of a gene index for cattle

JOURNAL

Genome Res. 11 (4), 626-630 (2001)

21180013

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -mnscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACACATGACAT

BACKWARD: GTTTTCCAGTCACGACG

Plate: 94 row: E column: 12

Seq primer: ATTTAGGTACATGATG.

Location/Qualifiers

1..549

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 1BOV"

/cissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site:1: XbaI; Site:2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 145 a 144 c 121 g 139 t

ORIGIN

alignment_scores: 888.00 Length: 183

Ratio: 5.078 Gaps: 1

Percent Similarity: 95.628 Percent Identity: 89.617

alignment_block:

US-09-171-553b-10 x AM657531 ..

Align seg 1/1 to: AM657531 from: 1 to: 549

290 SerProGlyValProValysThrGlyGlnArgLeuPheSerLeuIleG1 306

|||||

3 TCCCCCGGGTTCCTTAAAGACAGACAGACGCTTCAGCTCATCA 52

|||||

306 nGlyAlaPheGlnAlaIleAsnSerThrProAspAlaThrSerSerC 323

|||||

529 LeuGlnAsnArgArgGlyLeuAspLeuPheLeuArgGlyGlyLeu 545
 683 TTTCAAAACAGAGAGGTTAGATTATCTTTTAAAGGAGGAGGACT 732
 545 uCysAlaAlaLeuLysGlyLysCysPheTyrAlaAspHisSerGlyA 562
 733 ATGTGCGAGCTTAAAGAGAAATGCTTCTATGACACACAGCGGAA 782
 562 laileArgAspSerMetAsnLysLeuArgLysLeuGlnArgArg 578
 783 TAGTCAGAGATTCTATGCAAAATTGAGGAGAACTGGAAGGGA... 829
 579 ArgGluArgGluAlaAspGlnGlyTyrPheGlnGlyTyrPheAsnArgSe 595
 830TGGTTCAGAGGTGTGTACTCTCCTAATC 855
 595 rProTyrMetThrThrLeuLeuSer 603
 856 TCCATGATGACCACTTACTTCT 880

seq_name: gb_gss:BH122287

Seq documentation block: 774 bp DNA GSS 19-JUL-2001
 LOCUS BH122287
 DEFINITION RPI-24-288P13.TV RPI-24 Mus musculus genomic clone RPI-24-288P13
 , DNA sequence.
 ACCESION BH122287
 VERSION BH122287.1 GI:14965799
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 774)
 Zhao, S., Nieman, W., Malek, J., Shatsman, S., Akiret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Georegeorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPI-24
 Unpublished (1999)
 Other GSSs: RPI-24-288P13.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 288 row: P column: 13
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Source
 Location/Qualifiers
 1..774
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-24-288P13"
 /clone_id="RPI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1. Site 1: BamHI. Site 2: BamHI.
 RPI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."
 BASE COUNT 221 a 162 c 172 g 219 t
 ORIGIN

alignment_scores:
 Quality: 843.50 Length: 262
 Ratio: 3.749 Gaps: 2
 Percent Similarity: 85.878 Percent Identity: 61.832

alignment_block:
 us-09-171-553b-10 x BH122287

Align seg 1/1 to: BH122287 from: 1 to: 774

355 TrrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyTh 371
 2 TTTGGGGGAGAAACAAAGTTGACTCTGGCAGCACTTCAGAGAGAGCGCT 51
 371 rCysAlaGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThr 388
 52 TTTGTTGGGCCGACCTCAGATTAAGGACACCTCTGTAATCAGACC 101
 388 alValTyrGlnGlnAlaSerGlnAsnGlnTyrLeuValProGlyTyrAsn 404
 102 AGAATCATCAGCTCAGCAAAAGTGTGATCTAGTGGCCCCCTAGAC 151
 405 ArgTrrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerVa 421
 152 ACAGTGTGGGCTTGCATACCGGTCTCACCCTTGTGTGTGTGTGTGT 201
 421 lPheAsnGlnSerLysAspLeuCysValMetValGlnIleValProArg 438
 202 TTTTAATAGTTCACAAAGATTTCTGATTTGGTTCAGCTTATTCCTAGC 251
 438 alTyrTyrHisProGlnGluValValLeuAspLysIleTyrAspTyrArg 454
 252 TCTGTATCATGATGATGATGAGCGCTTTTACACAAATTTGACATCGG 301
 455 AsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeu 471
 302 ..CGCGGAGAGAGAGAGCCGTTACCTTACCTTGGCAGCTTATTTAG 348
 471 yLeuGlyThrAlaValAlaGlyThrGlyThrAlaAlaLeuIleThrG 488
 349 ATTGAGAGTAGCGGCTGAGTAGTACAGAACCGCTTATTTATTAAGA 398
 488 lProGlnGlnLeuGlnLysGlyLeuGlyLysLeuHisAlaAlaMetThr 504
 399 CCCCCCAATACTATGAA.....CACTACGTCGACGCTATGAT 436
 505 GlnAspLeuArgAlaLeuLysGlnSerValSerAsnLeuGlnLysSer 521
 437 GTTGATCTTGAACATATGAGACAGTCTATTAACCAATTAAGAAATCT 486
 521 uThrSerLeuSerGlnValValLeuGlnAsnArgArgGlyLeuAspLeu 538
 487 AACTTCCTCTCCGAAGTGTGTGTACAAATTAAGAGGAGTTGACTTAT 556
 538 eupheLeuArgGlnGlyLysLeuCysAlaAlaLeuLysGlnLysCys 554
 537 TATTCCTTAAAGAACGAGACTCTGTCTCTCTTAAAGAGAAATGTGGA 586
 555 PheTyrAlaAspHisSerGlyAlaIleArgAspSerMetAsnLysLeu 571
 587 TTTTATTTGACCATTCATGATTAAGATTAAGATTTATGAGCCCAACTT 636
 571 gLysLeuGlnGluArgArgArgGlnArgGlnAlaAspGlnGlyTrrp 588
 637 AGAAGCGCTAGATATAGTGAAGAGAAATAGAAACCAAGAGATGCT 686
 588 heGlnGlyTrrpPheAsnArgSerProTrrpMetThrThrLeuSerAla 604
 687 TCGAAGCTGGGTTAATTAAGTCCCTGGCTACACCTCTTCTCTCCACC 736
 605 LeuThrGlyProLeuValValLeuLeuLeuLeu 616

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (pages 1 to 845)	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	NIH-MGC http://mgc.ncl.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	unpublished (1999)
	Contact: Robert Strausberg	Ph. D.		

Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MCC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LRAM1181 row: f column: 10
 high quality sequence stop: 774.

FEATURES	Location/Qualifiers
source	1. .845

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="TMAG:5067273"
/clone_1b="NCL-CGAP-Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: Lung; Vector: pCMV-Sport6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT
ORIGIN
225 a 188 c 216 g 215 t
others

```

alignment_scores:	
Quality:	787.00
Ratio:	3.422
Percent Similarity:	79.861
	Percent Identity: 58.333
	Gaps: 8
	Length: 288

```
alignment_block:
US-09-171-553B-10 x BI151181 .
```

Align seg 1/1 to: B1151181 from: 1 to: 845

[illegible]

349 TTCATCTCATGATTT.TGTGTTCATGATCCACCTGTTACCCGGGTATTT 397
 440 ythiisProGluGluValValLeuaspGluTyrAspTyrArgTyrAsnArg 456
 398 ATCACCCTGCATCCAGTTTA.....GAGAAACCTTCCTGGCCGGCG 441
 457 ProTyrArgGluProValSerLeuThrLeuAlaValMetLeuGluLeuG 473
 442 TCAAAAAGGAGCAATTACTTAACTTACCTGGCTGCATCTCATGGGAATAG 491
 473 yThrAlaValGluValGluThrGluThrAlaAlaLeuIleHrGlyProG 490
 492 TATGGACGTAGAGAGTGGGACGGAGGTGTCACTTGTATGAAAGGAAC 541
 490 IGLIleuGluIuLysGluLeuGluLeuHnHsAlaIleMetThrGluAsp 506
 542 AG.....GGAAATCAGTCTTTGAGGGATGCTGTCAATGAGAC 579
 507 LeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSe 523
 580 CTAGAGCGCAATAGAGAACTCATCTTACGACGCTTTGAAAACCTTTGAC 629
 523 rLeuSerGluValValLeuGlnAsnAlaArgGlyLeuAspLeuLeuPhe 539
 630 CCTGTCTAGGTAGTTTACAGAAACAGGAAGGTGTTCATTTGTGTTC 679
 540 LeuArgGluGluGlyLeuGlyAlaAlaLeuLysGluGluGlyScySphET 556
 680 CTACAGGAAGGAGGACGTGTGTGCTGCCCTTAAGAAAGATGCTGCTTCT 729
 556 yValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgly 572
 730 ATGCAGATCATACAGGAATAGATTAGAAGACTTATCCCAAACTGAGAG 779
 572 slvs...LeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTTP 588
 780 AAGGATTCGAGCGCAGGAGACCGGAGACGGATGCTTCACGGGGTGGTG 829
 588 heGluGly 590
 830 AGTCCGG 837
 seq_name: gb_gss:A2850764
 seq_documentation_block: 641 bp DNA GSS 21-FEB-2001
 LOCUS A2850764 2M0152C18 Mouse 10kb plasmid U06CM library Mus musculus genomic
 DEFINITION 2M0152C18 Mouse 10kb plasmid U06CM library Mus musculus genomic
 clone U06C2M0152C18 R, DNA sequence.
 ACCESSION A2850764
 VERSION A2850764.1 GI:13036088
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 641)
 Dunn,D., Aoyagi,A., Barber,M., Beacom,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingley,A., von Niederhausern,A.
 and Wright,D.,Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 JOURNAL COMMENT
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid p1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

[illegible]

```
alignment_block:
US-09-171-553B-10 x AZ813042/rev
```

Align seg 1/1 to reverse of: AZ813042 from: 1 to: 625

```

408 AAcysaantnrglyLeuthrProxysValserThrServaLPhasngl 422
409 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 441
607 GCCTGCATTAACGGGCTCACTCCCTGGTGCTCACTGCTGTTTAAATG 555
424 nserLysaspLeucysValmetValGlnIleValProArgValIlyuTh 444
425 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 457
557 TTTCAAAAGATTTCTGCATTTTTGGTTCAGCTTTATCCGAGACCTGTGATC 508
441 iSPrgLuglValIleuAsngLdTyAspIlyrArgTyAsnArgPro 457
442 || :|||: |||||:|||||:|||||:|||||:|||||:||||| 507
ATGATGATAGCTCTTTTATAGATAAATTTGAACATCGGATC...CGCTGG 461
458 LysArgGluProValSerLeuThrLeuAlaValmetLeuGlyLeuGlyTh 474
460 AAAAGAGAACCCCGTTACTTAACCTTGGGAGTTCATTAAGATTTGGAGCT 4111
474 rAlaValGlyValGlyThGlyThrAlaAlaLeuIleThrGlyProGng 491
410 AGCAGCTGGAGGTAGGTACAGAGAACCCGCTCCCTTAATTAAAGACCCCAAT 361
491 ILeuGluIlyLysGlyLeuGlyIleuHnIsaAlaIleMetThGluAspLeu 507
360 ACTATGAA.....GAACTAGTGCGAGCTATGATATGAACTT 323
508 ArgAlaLeuLysGluSerValSerAsnLeuGluIleuSerLeuThrSerLe 524
322 AGAAGCTTAAGAACAGCTCTTAACCAAAATTAGAACATCTTAACTTCCCT 273
524 uSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuIleuAla 541
272 GTCTGAAGTGGCTCTCAAAATGAAGGAGATTAGACTTTATTCCTTA 223
541 rGgluglGlyLeuCysAlaAlaIleuLysGluIleuCysGlyPheTyVal 557
222 AAGAGAGAGACCTCTGCTGCTTAAAGAGAGATCTGTTTTATGTT 173
558 AsphHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysIle 574
172 GACCATTCAGAGATATCAAAAGATCTATGGCTAAACTTAGAACAGCTT 123
574 uGluArgArgArgGluIleuGluAlaAspGlnGlyTyrPheGluGlyT 591

```

122 AGATATACGTAAAGAGGAAGAGAGCCACAAAGAGATGTTCCAAACT 73
591 rppheasnaargserProtrpmetlrhrThrLeuLeuSerAlaLeuThrIly 60
72 GGTTAAATTAAGTCCCTTGCTGCTACACACTCTCTCCATATAGACGA 23
608 ProLeuValValLeuLeu 614
22 CCTTAATTACACTTAATGCTT 2

seq_name:	gb_est2:BI106669				
seq_documentation_block:		921 bp	mrna	EST	26-JUN-2001
LOCUS	BI106669				
DEFINITION	609293272f1 NC1-CGAP_Lu29	Mus	musculus	cdna	clone IMAGE:5038336 5'
ACCESSION	BI106669				
VERSION	BI106669.1	GI:14557562			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				

REFERENCE AUTHORS TITLE JOURNAL COMMENT	DATE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus 1 (bases 1 to 921) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D.	1999

Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM11105 row: p column: 17
High quality sequence stop: 805.

FEATURES	SOURCE
Location/Qualifiers	1. . 921
/organism="Mus musculus"	
/strain="C2ECH II (fetal)"	
/db_xref="taxon:10090"	
/clone="IMAGE:5038336"	
/clone_id="NCI_CGAP_Lu29"	
/tissue_type="spontaneous tumor, metastatic to mammary stem cell origin."	
/lab_host="DH10B"	
/note="Organ: Lung; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"	
BASE COUNT	249 a
ORIGIN	211 c 229 g 232 t

```
alignment_scores:
    Quality: 757.00
    Ratio: 3.181
    Gaps: 1d
Percent Similarity: 78.289
Percent Identity: 53.944
```

```
alignment_block:
US-09-171-553B-10 x BI106669
```

Align seg 1/1 to: B1106669 from: 1 to: 921

```

299 G|A|T|G|L|E|P|h|S|e|r|I|U|l|e|C|i|N|g|I|A|P|h|E|i|N|A|I|A|L|e|A|S|e|R|h   315
      |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
4    C|A|G|G|A|T|G|T|T|A|C|C|T|A|G|G|A|G|A|C|C|T|T|A|T|G|C|C|T|T|A|C|A|G|A|C   53
      |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
315 r|A|S|P|r|O|S|P|A|I|h|S|e|r|S|e|r|C|y|T|r|P|L|U|C|y|S|L|e|U|S|e|r|L|Y|P|r|P   332
      |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|

```

```

54  TATCCAGGCTACTGAGGACTGCTGGCTATGCTGCTGGGCTCCG 103
332  rctTyrGluGlyMetAlaLysGluArgLysPheAsnValThrIysGlu 348
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104  CTATTATTAGAGGAAATCGCTTCATGGAATTTCAACAGAACCGACG 153
349  HisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrG1 365
      ||| :||: ||||| ||||| ||||| ||||| ||||| |||||
154  CAT...ACTTCCTGCTCTGGGGTACAGACAAAAACAGACCTGACGTA 200
365  ValSerGlyLys.....GlyThrCysLeuGlyLysAlaPropSerH 380
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201  ACTATCCGCGAGGAAATCCAGCTCTGTATAGATCCCGACCTTCCACTC 250
380  iGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsn 396
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251  ACAAACACCTATGCGGACAAATTCAGTCCGCTCCAGAACGACAGCTGAT 300
397  GlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLe 413
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301  TACTATCTGTACTCTCCCGGTGAGTGGGCTGCTGCAATACAGACT 350
413  uhrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysV 430
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351  TACTCCCTGTATCACTAAAGTTTAAATTCATCATGATTTTCGCTG 400
430  aMetValGlnIleValProArgValTyrTyrHisProGluGluValAl 446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401  TATATATCCAGCTGTACCCGCTGTATATATACACCTGCATCCAGTTTA 450
447  LeuAspGluTyrAspTyrArgTyrAsnArgProLysArgLysProValS 463
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
451  GAAGACAGCTATGCGGCGCG.....CGGTCAAAAAGAAACCAATTAC 494
463  rleuThrIleuAlaValMetLeuGlyLeuGlyThrAlaValGlyValG 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495  TTTAACCCCTGCTGCATTCATGGAATAGTATGAGCTAGAGAGTGGGA 544
480  hGlyThrAlaAlaLeu..Ile..ThrGlyProGlnGlnLeuGlyLysG 496
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545  CGGGAGCTGACGCTTTCGATAGAACGGAAGACG.....GGAA 582
496  euGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu 512
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
583  TTCAGTCTTTGAGGATGCTGTCAATGAAGACCTAGCGCAATAGAGAG 632
513  SerValSerAsnLeuGlnGluSerLeuThrSerLeuSerGluValLe 529
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
633  TCCATTGACGCTTTAGAAAAATCTTGACCTCCCTGCTGAGTGTGTTT 682
529  uGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyLeuC 546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
683  ACGAAGACGAGAGGCTTTGATTCGTCGCTCAAAAGACAGACGCTGTT 732
546  ysaAlaLeuLysGluGlyCysCysPheTyr..ValAspHisSerGly..A 562
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
733  GTGCTGCCCTTAAGAGAGTGCCTCTATTTCAGATCAACAGAGAT 782
562  laIlaArgAspSer..MetAsnLysLeuArgLysLysLeuGluArg.. 577
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
783  TAGTTAGAGACTCTATAGCAACTGACAGAAACGATTTCAAGGAGG 832
578  ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsn 594
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
833  AACACGGGACCGGATGCCCAACGGGGTAGCTGAGTCCGGTTTGAATCA 882
594  Tg 594
      ||
883  GA 884

```

seq_name: gb_gss:AZ356870

seq_documentation_block:

```

LOCUS      AZ356870      644 bp      DNA
DEFINITION 1M0098F09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
ACCESSION  clone UUGC1M0098F09 F, DNA sequence.
VERSION    AZ356870
KEYWORDS   AZ356870.1 GI:10470556
SOURCE     GSS.
ORGANISM   house mouse.
            Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 644)
            Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
JOURNAL    Contact: Robert B. Weiss
COMMENT    University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert length: 10000 Std Error: 0.00
            Plate: 0098 Row: F Column: 09
            Seq primer: CGTTGTAACGACGCGCCAGT
            Class: plasmid ends
            High quality sequence stop: 644.
FEATURES   location/Qualifiers
            source          1..644
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="UUGC1M0098F09"
                        /clone_1ib="Mouse 10kb plasmid UUGC1M library"
                        /sex="Male"
                        /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                        /note="Vector: PWD42nv; Purified genomic DNA from M.
                        musculus C57BL/6J (male) was obtained from the Jackson
                        Laboratory Mouse DNA Resource
                        (http://www.jax.org/resources/documents/dnares/). The DNA
                        was hydrodynamically sheared by repeated passage through a
                        0.005 inch orifice at constant velocity. The sheared DNA
                        was blunt end-repaired with T4 DNA polymerase and T4
                        polynucleotide kinase. Adaptor oligonucleotides were
                        ligated to the blunt ends in high molar excess. The
                        adaptor DNA was purified and size-selected for a 9.5 to
                        10.5 kb range using preparative agarose gel
                        electrophoresis. Vector DNA was prepared from a derivative
                        of PWD42 (q14732114|gb|AF129072.1), a copy number
                        inducible derivative of plasmid R1. The vector was ligated
                        with adaptors complementary to the insert adaptors and
                        purified. The sheared, adaptor mouse DNA was annealed to
                        adaptor vector DNA, and transformed into
                        chemically-competent E. coli XL10-Gold (Stratagene) cells
                        and selected for ampicillin resistance."
BASE COUNT 185 a 136 c 133 g 190 t
ORIGIN

```

```

alignment_scores:
  Quality: 734.50      Length: 215
  Ratio: 3.866         Gaps: 2
  Percent Similarity: 88.372      Percent Identity: 66.512

```

alignment_block:

US-09-171-553B-10 x AZ356870/rev ..

Align seq 1/1 to reverse of: AZ356870 from: 1 to: 644


```

423 ACTACGTGACGTATGATGTGATCTTGAAGACTATAGACAGCTCTATATA 374
515 eTAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGluAsn 531
373 CCAATTTAGAGAAATCTTAACTTCCCTGTCGAAAGTGTCTACAGAAAT 324
533 ArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyLeuCysAlaI 548
333 AGAAGGGGATTAAGACTATATCTTAAAGAGAGAGACTCTGTCTGC 274
548 aLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArg 565
273 CTTAAAGAAAGATGTGTTTTTATGTGACCATTCAGAGCTAATCAAG 224
565 sPSeMetAsnLysLeuArgLysLysLeuGluArgArgArgGluArg 581
223 ATTCTATGGCCAACTTAGAGAGACGCTAGATATACGTAAGAGAGAAAGA 174
582 GluAlaAspGlnGlyTyrPheGluGluGlyTyrPheAsnArgSerProTyrPme 598
173 GAAAGCCAAAGAGATGTTGCGAAAGCTGTTTAAATAGTCCCTTGCT 124
598 tThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeu 615
123 CACCACTCTCTCTCCACCATAGCAGGACCTTTATTAACCTTATAGCTTT 74
615 eLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArg 631
73 TCCTTACTTTGGCCCTCGCATCTTATATAGTTAGTTAGCTTTATTAGA 24
632 GluArgValSer 635
23 GAAAGATTAAT 12
seq_name: gb_gss:AZ430249
seq_documentation_block:
LOCUS AZ430249 591 bp DNA GSS 03-OCT-2000
DEFINITION 1M0214G11R Mouse 10kb plasmid UNGC1M library Mus musculus genomic
clone UNGC1M0214G11 R, DNA sequence.
ACCESSION AZ430249
VERSION AZ430249.1 GI:10554262
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 591)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Rellly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0214 row: 6 column: 11
Seq primer: CACACGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 591.
Location/Qualifiers
FEATURES
source
1..591
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```

```

/clone="UNG1M0214G11"
/clone.lib="Mouse 10kb plasmid UNGC1M library"
/sex="Male"
/lab.host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (911473211419b|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor DNA was annealed to
adaptor DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

```

```

BASE COUNT 165 a 125 c 115 g 186 t
ORIGIN

```

```

alignment_scores:
Quality: 729.00 Length: 194
Ratio: 4.096 Gaps: 1
Percent Similarity: 91.753 Percent Identity: 71.134

```

```

alignment_block:

```

```

US-09-171-553b-10 x AZ430249/rev

```

```

Align seq 1/1 to reverse of: AZ430249 from: 1 to: 591

```

```

456 ArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLe 472
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588 CGCGGGAAGAAAGAGAACCCGTTACTTACCTTGCGAGTTTATTAAGATT 539
472 uGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyP 489
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
538 GGGAGTAGACAGCTGAGTAGTAGACAGAACCCGCTGCTTAATTAACACCC 489
489 rGtGlnGlnLeuGluLysGlyLeuGluGluLeuHisAlaAlaMetThrGlu 505
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
488 CCCAATCTATGAA.....GAGCTACGTCAGCTATGATGATT 451
506 AspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuTh 522
450 GATCTTAGAACCATAGACAGCTCTATTAACCAATTAAGAAATCTTTAAC 401
522 rSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuP 539
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 TTCCCTATCCGAAGCTGTGCTGCAAAATGAAGAGGGGTTAGACTTATTAT 351
539 heLeuArgGluGlyGlyLeuCysAlaLeuLysGluGluCysCysPhe 555
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 TCCTTAAGAGAGAGAGACTCTGTGCTCCCTTAAGAAAGAAATGTTGTTT 301
556 TyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgly 572
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 TATGTGACCATTCAGAGATTAATCAAGATTCATGCTTAATCAATTAAGA 251
572 sLysLeuGluArgArgArgGluArgGluAlaAspGlnGlyTyrPheG 589
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 AGCCCTAGATATACGTAAGAAAGAGAAAGAACCCACAGAGATGTTTG 201
589 uGlyTyrPheAsnArgSerProTyrPmeThrThrLeuLeuSerAlaLeu 605
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 AAGCGTGTTTAAATAGTCCCTTGCTGTCACGACTCTCCCTCCACACTATA 151

```



```

454 yrasnmgProLysaRGILpProLiseriutInLeuIaValMetIeu 4
      ::::| | | | | | | | | | | | | | | | | | | | | |
95 GCCGGCGGTCAAAAAGAACCAATTACTTTAACCTGGCTGCATTCATG 144

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

195 AGAAGACAG.....GGAATTCAGTCTTGAGGAGATGCTGTCA 232

Location/Qualifiers
1. .673

```
/organism="Mus musculus"
/strain="C57BL/6J"
```

```
/db_xref="taxon:10090"
```

```
/clone="UUGC1M0068F24"
/clone 1ib="Mouse 10kb r
```

```
clone_id = mouse_id
sex = "Male"
```

```
/lab_host="E. coli strain
/notes="Vector: pMD19uv.
```

musculus C57BL/6J (male),
/MOE- vector: FMD42IV,

Laboratory Mouse DNA Res

was hydrodynamically she

0.005 inch orifice at co

was blunt end-repaired with

ligated to the blunt end

10.5 kb range using prep

electrophoresis. Vector

Inducible derivative of
ot pWD42 (g1|4/32|14|gb

with adaptors complement

purified. The sheared, adapted vector DNA, and

chemically-competent E.

a and selected for amplification 120 c 167 g 18

100

Length	Genotype
721.50	721.50
3.607	3.607

Percent Identical	Percent Identical
87.719	87.719

AZ337993 ..

AZ3337993 from: 1 to: 67

4
3
4
4
5
5
6
8
8
9
9
10
10
11
11
12
12
13
13
14
14
15
15
16
16
17
17
18
18
19
19
20
20
21
21
22
22
23
23
24
24
25
25
26
26
27
27
28
28
29
29
30
30
31
31
32
32
33
33
34
34
35
35
36
36
37
37
38
38
39
39
40
40
41
41
42
42
43
43
44
44
45
45
46
46
47
47
48
48
49
49
50
50
51
51
52
52
53
53
54
54
55
55
56
56
57
57
58
58
59
59
60
60
61
61
62
62
63
63
64
64
65
65
66
66
67
67
68
68
69
69
70
70
71
71
72
72
73
73
74
74
75
75
76
76
77
77
78
78
79
79
80
80
81
81
82
82
83
83
84
84
85
85
86
86
87
87
88
88
89
89
90
90
91
91
92
92
93
93
94
94
95
95
96
96
97
97
98
98
99
99
100
100
101
101
102
102
103
103
104
104
105
105
106
106
107
107
108
108
109
109
110
110
111
111
112
112
113
113
114
114
115
115
116
116
117
117
118
118
119
119
120
120
121
121
122
122
123
123
124
124
125
125
126
126
127
127
128
128
129
129
130
130
131
131
132
132
133
133
134
134
135
135
136
136
137
137
138
138
139
139
140
140
141
141
142
142
143
143
144
144
145
145
146
146
147
147
148
148
149
149
150
150
151
151
152
152
153
153
154
154
155
155
156
156
157
157
158
158
159
159
160
160
161
161
162
162
163
163
164
164
165
165
166
166
167
167
168
168
169
169
170
170
171
171
172
172
173
173
174
174
175
175
176
176
177
177
178
178
179
179
180
180
181
181
182
182
183
183
184
184
185
185
186
186
187
187
188
188
189
189
190
190
191
191
192
192
193
193
194
194
195
195
196
196
197
197
198
198
199
199
200
200
201
201
202
202
203
203
204
204
205
205
206
206
207
207
208
208
209
209
210
210
211
211
212
212
213
213
214
214
215
215
216
216
217
217
218
218
219
219
220
220
221
221
222
222
223
223
224
224
225
225
226
226
227
227
228
228
229
229
230
230
231
231
232
232
233
233
234
234
235
235
236
236
237
237
238
238
239
239
240
240
241
241
242
242
243
243
244
244
245
245
246
246
247
247
248
248
249
249
250
250
251
251
252
252
253
253
254
254
255
255
256
256
257
257
258
258
259
259
260
260
261
261
262
262
263
263
264
264
265
265
266
266
267
267
268
268
269
269
270
270
271
271
272
272
273
273
274
274
275
275
276
276
277
277
278
278
279
279
280
280
281
281
282
282
283
283
284
284
285
285
286
286
287
287
288
288
289
289
290
290
291
291
292
292
293
293
294
294
295
295
296
296
297
297
298
298
299
299
300
300
301
301
302
302
303
303
304
304
305
305
306
306
307
307
308
308
309
309
310
310
311
311
312
312
313
313
314
314
315
315
316
316
317
317
318
318
319
319
320
320
321
321
322
322
323
323
324
324
325
325
326
326
327
327
328
328
329
329
330
330
331
331
332
332
333
333
334
334
335
335
336
336
337
337
338
338
339
339
340
340
341
341
342
342
343
343
344
344
345
345
346
346
347
347
348
348
349
349
350
350
351
351
352
352
353
353
354
354
355
355
356
356
357
357
358
358
359
359
360
360
361
361
362
362
363
363
364
364
365
365
366
366
367
367
368
368
369
369
370
370
371
371
372
372
373
373
374
374
375
375
376
376
377
377
378
378
379
379
380
380
381
381
382
382
383
383
384
384
385
385
386
386
387
387
388
388
389
389
390
390
391
391
392
392
393
393
394
394
395
395
396
396
397
397
398
398
399
399
400
400
401
401
402
402
403
403
404
404
405
405
406
406
407
407
408
408
409
409
410
410
411
411
412
412
413
413
414
414
415
415
416
416
417
417
418
418
419
419
420
420
421
421
422
422
423
423
424
424
425
425
426
426
427
427
428
428
429
429
430
430
431
431
432
432
433
433
4

msrLysaspLeucysValMetVal...

ATCTCATGATTTTGTGTCATGATC

```
isProgluValValLeuAspG
```

ACCCCTGCATCCAGTTTA.....GA

1

University of Utah Genome Center
University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunne@genetics.utah.edu

Insert length: 10000 Std Error: 0.00

Plate: 0081 row: N column: 06

Seq primer: CACACAGAAACAGCTATGACAC

Class: plasmid ends

High quality sequence stop: 632.

FEATURES

SOURCE

1. 632
Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUCG2M0081N06"

/clone_lib="Mouse 10kb plasmid UUCG1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pMD42 (g14732114[gblAF129072.1]), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 203 a 115 c 138 g 176 t
ORIGIN

alignment_scores:

Quality: 713.50

Length: 199

Ratio: 3.920

Gaps: 2

Percent Similarity: 91.457

Percent Identity: 68.844

alignment_block:

US-09-171-553B-10 x AZ814037

Align seg 1/1 to: AZ814037 from: 1 to: 632

```

459 ArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAl 475
|||||
3 AGAGAGACCCGTTACTTACTTGGCAGTTCTATTAGGATTGGAGTAGT 52
475 aValGlyValGlyThrGlyThrAlaValMetLeuGlyProGlyngInL 492
|||||
53 GGCTGAGTAGTACAGAGACTGCTGCTTATTAAGACCCCAATACT 102
492 eugGluValGlyLeuGlyLeuHisAlaValMetThrGluAspLeuArg 508
|||||
103 ATGAA.....GAACTACGTGACGTATGATGTTAGCTTAGA 140
509 AlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSe 525
|||||
141 ACTATGACAGCTCTATACCAATATAGAAATCTTTACTTCCTGTC 190
525 rGluValValLeuGluInsAsnArgArgGlyLeuAspLeuLeuPheLeuArg 542
|||||
191 CGAAGTGGTCTACAGATAGAAAGGAGATTAGCTTATTTCTTAAG 240
542 LuGlyGlyLeuGlyAlaLeuLysGluGlyCysPheTyrValAsp 558

```

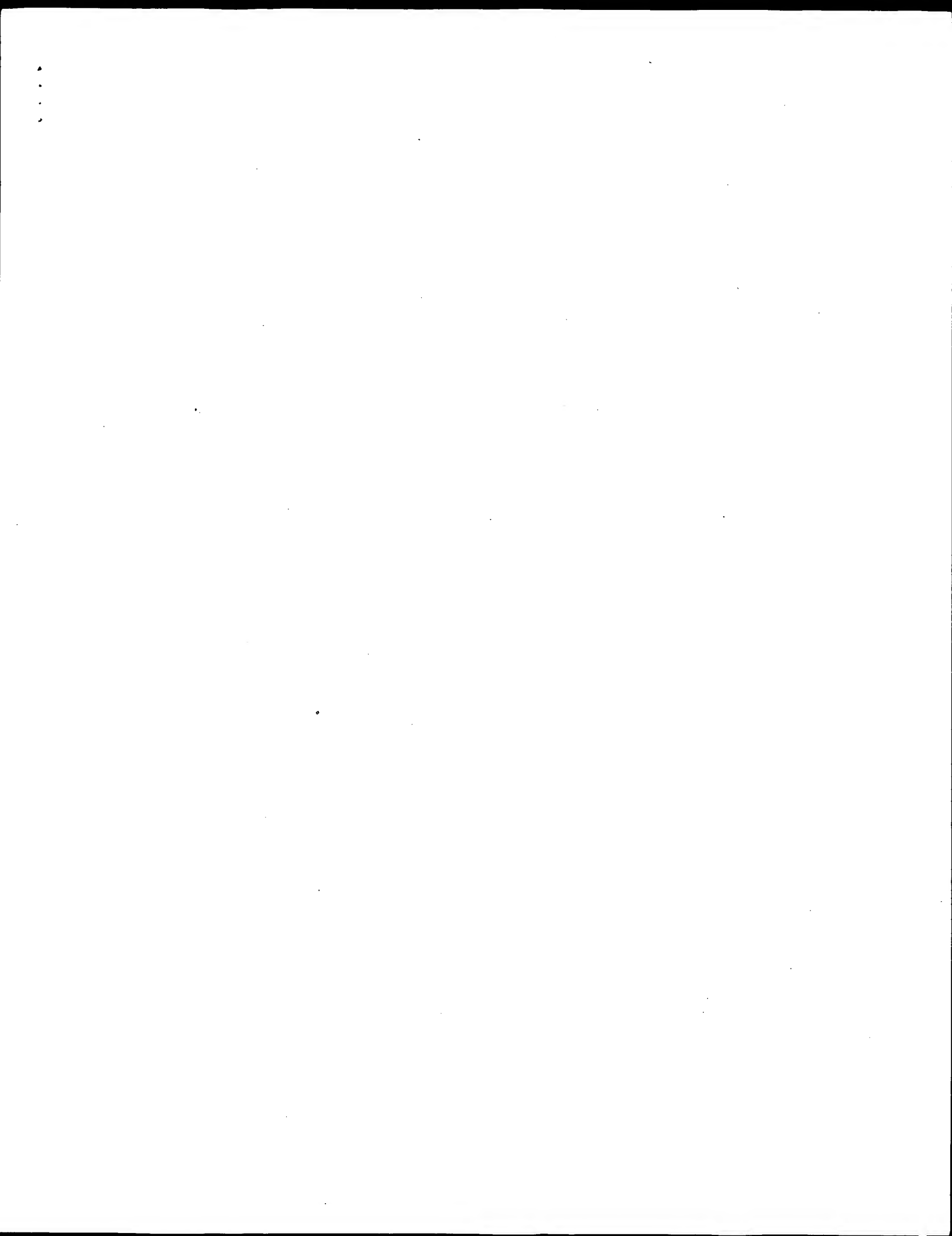
```

241 AAGGAGACTCTGTGCTGCCCTAAAGAGATGTTGTTTATGTTGAC 290
559 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuG 575
|||||
291 CATTCAAGAGATTAATCAAGATTCATGCGCCAACTTAGAAGCGCTAGA 340
575 uArgArgArgGluArgGluAlaAspGlyngLysTyrPheGlyGlyTyrP 592
|||||
341 TATACGTAAAGAGAAAGAGAAAGAGAAAGAGAGATGTTGAAAGCTG 390
592 heAsnArgSerProTyrMetThrThrLeuLeuSerAlaLeuThrGlyPro 608
|||||
391 TTATATAGTCCCTTGGCTGACCACTCTCTCCACATAGAGAGAGAGCT 440
609 LeuValValLeuLeuLeuLeuThrValGlyProCysLeuLeuAsnArg 625
|||||
441 TTATATACCTTATGCTTGTCTTACTTTGCTCCTGCAATCCTTATTA 490
625 gPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVal 642
|||||
491 GTTAGTGTGCTTTATTAGAAGAGATTAATGACAGTATAGTATGTCAC 540
642 euArgGlyngLysGlyLeuLeuSerGlyngLysGlyLeuThrAspLeu 657
|||||
541 TAAAGCAACAATATCGG...GTCCCTCAGAGGTTGAAACTGCTC 584

```


' Mon Feb 25 07:44:04 2002

us-09-171-553b-10.feb22std.rst



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:18 ; Search time 6550.43 Seconds
(without alignments)
32.809 Million cell updates/sec

Title: US-09-171-553b-13

Perfect score: 20

Sequence: 1 gatgctctctgccccttgg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estfun:
2: em_esthum:
3: em_estin:
4: em_estom:
5: em_estpl:
6: em_estba:
7: em_estro:
8: em_estov:
9: em_hlc:
10: gb_est1:
11: gb_est2:
12: gb_hlc:
13: gb_gss:
14: em_gss_fun:
15: em_gss_hum:
16: em_gss_inv:
17: em_gss_pln:
18: em_gss_pro:
19: em_gss_rtd:
20: em_gss_vrt:
21: em_gss_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	391	11	BF465634 UI-M-CGDP
2	18.4	92.0	429	11	AA797871 vW31h06.r
3	18.4	92.0	435	11	BF565386 UI-R-B01
4	18.4	92.0	444	11	AA472060 v97d05.r
5	18.4	92.0	488	11	BF719561 mab41e06.r
6	18.4	92.0	510	13	AA033199 m36d04.r
7	18.4	92.0	539	13	AA0512916 HS_5139.A
8	18.4	92.0	566	10	AA121599 zN78a06.r
9	18.4	92.0	571	10	BF097309 UI-R-B01
10	18.4	92.0	592	10	AM556515 L0269D06-
11	18.4	92.0	600	10	AM488316 UI-M-BH3
12	18.4	92.0	659	10	BE374286 601227645

C 13	18.4	92.0	690	11	BG075579	BG075579 H3149A02-
C 14	18.4	92.0	896	11	BG342613	BG342613 602374796
C 15	18.4	92.0	2171	12	AK019522	AK019522 Mus muscu
C 16	17.4	87.0	257	10	AM326534	AM326534 19216 MAR
C 17	17.4	87.0	397	10	AM489995	AM489995 UI-M-BH3-
C 18	17.4	87.0	455	10	AT013028	AT013028 EST207479
C 19	17.4	87.0	581	13	A2777840	A2777840 2M0012D10
C 20	17.4	87.0	607	11	W20084	W20084 zB40F01.r1
C 21	17.4	87.0	783	13	BH051324	BH051324 RPT-24-3
C 22	17	85.0	691	10	AT735229	AT735229 at08a01.x
C 23	17	85.0	1071	10	BE612516	BE612516 601452020
C 24	17	85.0	1787	11	BG256691	BG256691 602370880
C 25	16.8	84.0	190	11	T25876	T25876 ESTDIR3 CD3
C 26	16.8	84.0	219	10	B8362510	B8362510 B8362510
C 27	16.8	84.0	250	10	AA478651	AA478651 zV19c05.r
C 28	16.8	84.0	268	10	AV228430	AV228430 AV228430
C 29	16.8	84.0	422	13	AO902607	AO902607 LMAJFV1_1
C 30	16.8	84.0	454	11	BF549019	BF549019 UI-R-A0-a
C 31	16.8	84.0	473	10	AA170346	AA170346 ms88909.r
C 32	16.8	84.0	492	11	BF54951	BF54951 UI-R-B0-c
C 33	16.8	84.0	504	13	A2582584	A2582584 1M0376003
C 34	16.8	84.0	536	13	AO456957	AO456957 HS_5151_A
C 35	16.8	84.0	539	10	BE586605	BE586605 WHE0509_G
C 36	16.8	84.0	555	10	BE650111	BE650111 UI-M-BH3-
C 37	16.8	84.0	558	13	AO523881	AO523881 HS_5246_A
C 38	16.8	84.0	576	13	A2730616	A2730616 RPT-24-1
C 39	16.8	84.0	594	13	A2951110	A2951110 2M0215M20
C 40	16.8	84.0	602	10	A1102103	A1102103 EST211392
C 41	16.8	84.0	610	13	A2432720	A2432720 1M0218E17
C 42	16.8	84.0	610	13	A2859488	A2859488 2M0165D07
C 43	16.8	84.0	654	13	A2741186	A2741186 RPT-24-6
C 44	16.8	84.0	659	13	A2453887	A2453887 1M0255N03
C 45	16.8	84.0	658	13	A2457973	A2457973 1M0261A07

ALIGNMENTS

RESULT 1
BF465634 391 bp mRNA EST 04-DEC-2000
UI-M-CGDP-bq1-h-12-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
UI-M-CGDP-bq1-h-12-0-UI 3', mRNA sequence.

BF465634
BF465634.1 GI:11534817

ACCESION
VERSION
KEYWORDS
SOURCE
ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 391)

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

JOURNAL

MEDLINE

COMMENT

FEATURES

location/Qualifiers

Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 8990
Email: mestr@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENERICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA-No.

```

1. 429
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1245467"
/clone_id="Soares_mammary_gland_NbMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco

```

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="ur-R-B01-ajv-e-06-0-ur"
/clone_1ib="ur-R-B01"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/vector="pT73D-Pic (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The library
(ur-R-B01) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonatido, Lennon and Soares, Genome Research
6:791-806, 1996)"

```

BASE COUNT 105 a 121 c 115 g 94 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 11; Length 435;
Best Local Similarity 95.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 gatgctctcgccttg 20
|||||
Db 346 GATGCTCTCTCCCTTGG 365

RESULT 4

AA472060 444 bp mRNA EST 18-JUN-1997
LOCUS v977d05.t1 Soares_mammary_gland_NbMKG Mus musculus cDNA clone
DEFINITION IMAGE:873897 5' similar to TR:G190352 G190352 PEREGRIN.; mRNA
sequence.

ACCESSION AA472060

VERSION AA472060.1 GI:2200051

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 444)
Matta, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucada, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The Mashu-HMI Mouse EST Project
JOURNAL Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project
Washu-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:513377

Seq primer: -28m13 rev2 EF from Amersham.

FEATURES

source

Location/Qualifiers

1..444

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:873897"

/clone_lib="Soares_mammary_gland_NbMKG"

/sex="male"

/tissue_type="mammary gland"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer (5'
TGTTACCAATCTGAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT
T 3'); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia); digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT 104 a 116 c 127 g 97 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 444;
Best Local Similarity 95.0%; Pred. No. 4.4e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 gatgctctcgccttg 20
|||||
Db 301 GATGCTCTCTCCCTTGG 320

RESULT 5

BF719561 488 bp mRNA EST 03-JAN-2001
LOCUS mab41e06.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone
DEFINITION IMAGE:3972947 5' similar to SW:BR14_HUMAN P55201 PEREGRIN.; mRNA
sequence.

ACCESSION BF719561 GI:12020563

VERSION BF719561

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 488)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL),
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
image.llnl.gov/image/html/resources.shtml

MGI:1472979

Seq primer: -40RP from Gibco

High quality sequence stop: 460.

FEATURES

source

Location/Qualifiers

1..488

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:3972947"

/clone_lib="Soares_NMEBA_branchial_arch"

/tissue_type="branchial arches"

/dev_stage="embryo, 10.5 dpc"

/lab_host="DH10B (phage resistant)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer (5'
TGTTACCAATCTGAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT
3'); double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia); digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT 115 a 135 c 128 g 110 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 11; Length 488;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 gatgctctcgccttg 20
|||||
Db 425 GATGCTCTCTCCCTTGG 444

RESULT 6

AA033199 510 bp mRNA EST 22-AUG-1996
LOCUS m136d04.t1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
DEFINITION

TITLE JOURNAL MEDLINE COMMENT

Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnlni.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 439.

FEATURES

SOURCE

1. 566
/organism="Homo sapiens"
/db_xref="GDB:4596797"
/db_xref="taxon:9606"
/clone="IMAGE:564274"
/tissue="Stratagene NT2 neuronal precursor 937230"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (Kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG
3' -3' adaptor sequence: 5' CTCAGTGTGTGTGTGTGTGTGT 3'."

BASE COUNT 114 a 157 c 141 g 151 t 3 others
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 566;
Best Local Similarity 95.0%; Pred. No. 4.5e-02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gatggctctctgccccttg 20
|||||
Db 348 GAGGGCTCTCCTGCCCTTG 329

RESULT 9

BE097309/c
LOCUS BE097309 571 bp mRNA EST 12-JUN-2000
DEFINITION UI-R-B01-apv-g-12-0-UI.sl UI-R-B01 Rattus norvegicus cDNA clone
UI-R-B01-apv-g-12-0-UI 3', mRNA sequence.
ACCESSION BE097309
VERSION BE097309.1 GI:8488240
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 571)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the

normalized medulla library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

SOURCE

Location/Qualifiers
1. 571
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-B01-apv-g-12-0-UI"
/clone_lib="UI-R-B01"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The library
(UI-R-B01) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=UI-R-B01
TAG_TISSUE=medulla
TAG_SEQ=GAACCG"
BASE COUNT 143 a 152 c 134 g 141 t 1 others
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 571;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gatggctctctgccccttg 20
|||||
Db 528 GATGGCTCTCCTGCCCTTG 509

RESULT 10

AW556515/c
LOCUS AW556515 592 bp mRNA EST 31-AUG-2000
DEFINITION L0269D06-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
clone L0269D06 3', mRNA sequence.
ACCESSION AW556515
VERSION AW556515.1 GI:7201944
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 592)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE 20381348
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: L0269 row: D column: 06
Seq primer: -21M13 Forward
High quality sequence stop: 592
POLYA=Yes.
Location/Qualifiers
1. 592
/organism="Mus musculus"

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L0269006"
/clone_lib="NIA Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/notes="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an oligo(dT) primer [NotI
primer-adaptor from GibcoBRL]
15'-pGACTAGTCTAGATCCGAGCGCGCCCTTTTITTTT-3'] from
2.5µg of total RNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker L1-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Yulan Piao." 3 others
BASE COUNT      140 a 166 c 146 g 137 t
ORIGIN

```

```

Query Match      92.0%; Score 18.4; DB 10; Length 592;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1 gatggctctctgcctttg 20
      |||||
Db   511 GATGGCTCTCTCCCTTCCTTG 492

```

```

RESULT 11
AW488316/C
LOCUS      AW488316      600 bp      mRNA      EST      24-FEB-2000
DEFINITION UI-M-BH3-arp-b-06-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
ACCESSION  AW488316
VERSION    AW488316.1 GI:7058586
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 600)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            9704477
            Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mst@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized cerebellum library cDNA Library Preparation: M.B. Soares
            Lab Clone distribution: Researchers may obtain BMAP cDNA clones

```

from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward

POLYA-Yes.

```

FEATURES             Location/Qualifiers
     source            1..600
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-arp-b-06-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /notes="Vector: pT7f3D-Pac (Pharmacia) with a modified
                     polylinker: Site 1: Not I; Site 2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately derived from a mixture of individually tagged
                     normalized libraries from ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland, striatum,
                     hippocampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S1.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
                     (NIH_BMAP_M_S4) was constructed as follows: PCR amplified
                     cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                     NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                     was used as a driver in a hybridization with a pool of
                     the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                     libraries in the form of single-stranded circles. The
                     remaining single-stranded circles (subtracted library)
                     was purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B bacteria (Life Technologies) to generate the
                     NIH_BMAP_M_S4 library. This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB=NIH_BMAP_M_S4
                     TAG_TISSUE=cerebellum
                     TAG_SEQ=GACTC"
BASE COUNT      141 a 166 c 143 g 150 t
ORIGIN

```

```

Query Match      92.0%; Score 18.4; DB 10; Length 600;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1 gatggctctctgcctttg 20
      |||||
Db   528 GATGGCTCTCTCCCTTCCTTG 509

```

```

RESULT 12
BE374286
LOCUS      BE374286      659 bp      mRNA      EST      21-JUL-2000
DEFINITION 601227645F1 NCL_CGAP_Maml Mus musculus cDNA clone IMAGE:3585899 5',
            mRNA sequence.
ACCESSION  BE374286
VERSION    BE374286
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 659)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)

```


COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Cloned through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM8746 row: j column: 12
 High quality sequence stop: 586.
 Location/Qualifiers
 1. .659

FEATURES
source

/organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3585899"
 /clone_lib="NCI_CGAP_Mam1"
 /dev_stage="3 months, virgin"
 /tissue_type="tumor, biopsy sample"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT
ORIGIN

149 a 178 c 189 g 143 t
 Query Match 92.0%; Score 18.4; DB 10; Length 659;
 Best Local Similarity 95.0%; Pred. No. 4.5e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatggctctcctgccttg 20
 ||||| ||||| ||||| |||||
 Db 452 GATGGCTCTCTTCCTTTG 471

RESULT 13
BG075579/c

LOCUS BG075579 690 bp mRNA EST 26-JAN-2001
 DEFINITION H3149A02-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3149A02 3', mRNA sequence.
 ACCESSION BG075579
 VERSION BG075579.1 GI:12558148
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 690)
 AUTHORS T.S., Carter, M.G. and Ko, M.S.H.
 TITLE Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka
 Verification and initial annotation of NIA mouse 15K cDNA clone set
 Unpublished (2001)
 Other_ESTs: H3149A02-5
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdaei@gsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please
 visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
 Plate: H3149 row: A column: 02
 Seq primer: -21M13 Forward
 High quality sequence stop: 690
 POLYA=Yes.

Location/Qualifiers
 1. .690
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="H3149A02"

FEATURES
source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="H3149A02"

/clone_lib="NIA Mouse 15K cDNA Clone Set"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
 clone is among a rearranged set of 15,247 clones from 11
 embryo cDNA libraries (including preimplantation stage
 embryos from unfertilized egg to blastocyst, embryonic
 part of E7.5 embryos, extraembryonic part of E7.5 embryos
 , and E12.5 female mesonephros/gonad) and one newborn
 ovary cDNA library. Average insert size 1.5 kb. All
 source libraries are cloned unidirectionally with Oligo(dT
)-Not primers. References include: (1) Genome-wide
 expression profiling of mid-gestation placenta and embryo
 using a 15,000 mouse developmental cDNA microarray, 2000,
 Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
 Large-scale cDNA analysis reveals phased gene expression
 patterns during preimplantation mouse development, 2000,
 Development, 127: 1737-1749; (3) Genome-wide mapping of
 unselected transcripts from extraembryonic tissue of
 7.5-day mouse embryos reveals enrichment in the t-complex
 and under-representation on the X chromosome, 1998, Hum
 Mol Genet 7: 1967-1978."

BASE COUNT 158 a 197 c 170 g 165 t
 ORIGIN

Query Match 92.0%; Score 18.4; DB 11; Length 690;
 Best Local Similarity 95.0%; Pred. No. 4.5e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatggctctcctgccttg 20
 ||||| ||||| ||||| |||||
 Db 512 GATGGCTCTCTTCCTTTG 493

RESULT 14
BG342613

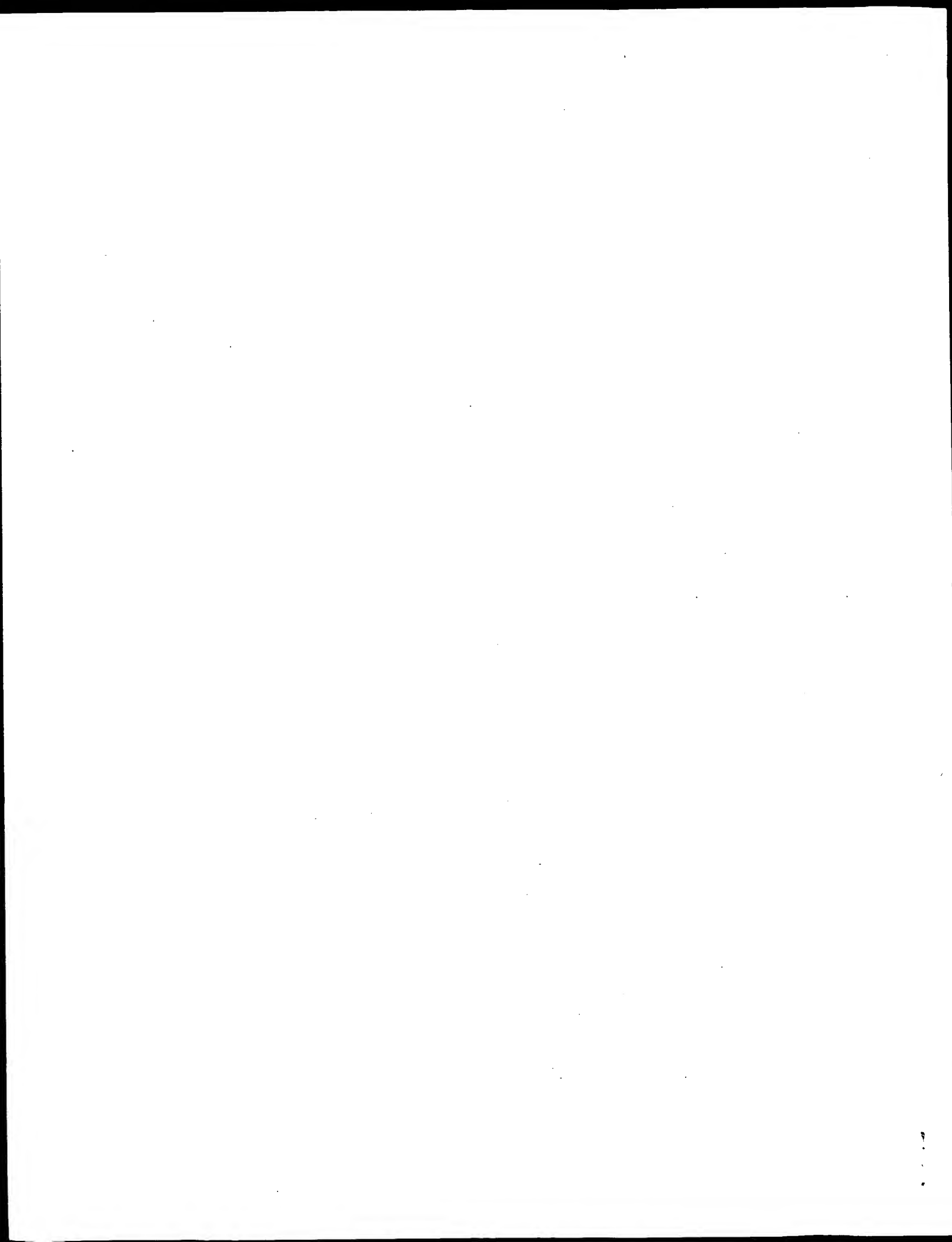
LOCUS BG342613 896 bp mRNA EST 27-FEB-2001
 DEFINITION 602374796f1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4482310 5',
 mRNA sequence.
 ACCESSION BG342613
 VERSION BG342613.1 GI:13149051
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 896)
 AUTHORS NTH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov

Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Cloned through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10318 row: p column: 23
 High quality sequence stop: 564.
 Location/Qualifiers
 1. .896

FEATURES
source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4482310"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:21 : Search time 6550.43 Seconds
(without alignments)
29.528 Million cell updates/sec

Title: US-09-171-553B-14
Perfect score: 18
Sequence: 1 ccacagtcgtacaccacg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	18	100.0	558	11	BI360879
c 2	17	94.4	323	13	AZ056887
c 3	17	94.4	559	13	AZ749422
c 4	17	94.4	655	13	BH097469
c 5	16.4	91.1	155	10	AW346676
c 6	16.4	91.1	166	10	AW346655
c 7	16.4	91.1	232	10	AV312947
c 8	16.4	91.1	254	10	AV289929
c 9	16.4	91.1	261	11	BI184980
c 10	16.4	91.1	290	13	AZ113096
c 11	16.4	91.1	294	11	BF713588
c 12	16.4	91.1	300	11	BI184279

c 13	16.4	91.1	370	13	AZ313323
c 14	16.4	91.1	379	13	AZ793176
c 15	16.4	91.1	395	13	AZ831760
c 16	16.4	91.1	425	13	AZ036915
c 17	16.4	91.1	427	13	AZ58978
c 18	16.4	91.1	432	13	AZ909325
c 19	16.4	91.1	458	11	BI185480
c 20	16.4	91.1	468	11	BI182727
c 21	16.4	91.1	468	11	BI186066
c 22	16.4	91.1	470	13	AZ387701
c 23	16.4	91.1	495	13	AZ955800
c 24	16.4	91.1	499	11	BF704354
c 25	16.4	91.1	513	10	AW786160
c 26	16.4	91.1	521	11	BI399852
c 27	16.4	91.1	552	13	AZ242483
c 28	16.4	91.1	553	13	AZ999118
c 29	16.4	91.1	579	11	BI183723
c 30	16.4	91.1	602	11	BF712162
c 31	16.4	91.1	607	13	BH088353
c 32	16.4	91.1	613	13	AZ454083
c 33	16.4	91.1	615	13	AZ385535
c 34	16.4	91.1	622	13	AZ115761
c 35	16.4	91.1	655	13	AZ836447
c 36	16.4	91.1	681	13	AZ115038
c 37	16.4	91.1	683	11	BG122508
c 38	16.4	91.1	767	13	BH125701
c 39	16.4	91.1	864	11	BF790922
c 40	16.4	91.1	1042	10	BE544440
c 41	16.4	91.1	2260	12	AK015210
c 42	16	88.9	686	11	BF501296
c 43	15.4	85.6	142	13	AZ601453
c 44	15.4	85.6	252	11	W66405
c 45	15.4	85.6	259	13	AZ308532

ALIGNMENTS

RESULT 1

BI360879/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI360879 558 bp mRNA EST 01-AUG-2001
388957 MARC 2Pig Sus scrofa cDNA 5', mRNA sequence.
BI360879
BI360879.1 GI:15056907

EST

Pig.

Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheraia; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 558)

Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,

Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.

and Keeler, J.W.

Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred

v0.980504 e. Vector identified by cross_match with the -mnscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACAGC

Plate: 142 row: B column: 4

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..558

/organism="Sus scrofa"

FEATURES
source

```

/db_xref="taxon:9823"
/clone_lib="MARC 2PIC"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      129 a 123 c 119 g 187 t
ORIGIN

Query Match      100.0%; Score 18; DB 11; Length 558;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 18
|||||
Db 178 CCACAGTCGTACACCACG 161

RESULT 2
LOCUS      AZ056887      323 bp      DNA      GSS      30-MAR-2000
DEFINITION RPCI-23-428K7.TJ RPCI-23 Mus musculus genomic clone RPCI-23-428K7,
DNA sequence.
ACCESSION  AZ056887
VERSION    AZ056887.1 GI:7348123
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 323)
AUTHORS    Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL    Unpublished (1999)
COMMENT    Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 428 row: K column: 7
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
source      1..323
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-428K7"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      83 a 59 c 72 g 109 t
ORIGIN

```

```

Query Match      94.4%; Score 17; DB 13; Length 323;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
|||||
Db 218 CCACAGTCGTACACCAC 234

RESULT 3
LOCUS      AZ749422      559 bp      DNA      GSS      25-JAN-2001
DEFINITION RPCI-24-115E19.TV RPCI-24 Mus musculus genomic clone RPCI-24-115E19
, DNA sequence.
ACCESSION  AZ749422
VERSION    AZ749422.1 GI:12534581
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 559)
AUTHORS    Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-24
JOURNAL    Unpublished (1999)
COMMENT    Other GSSs: RPCI-24-115E19.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC
Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 115 row: E column: 19
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
source      1..559
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-115E19"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      133 a 125 c 137 g 164 t
ORIGIN

Query Match      94.4%; Score 17; DB 13; Length 559;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
|||||
Db 411 CCACAGTCGTACACCAC 395

RESULT 4
BH097469

```

LOCUS
DEFINITION RPCI-24-315L22.TV RPCI-24 Mus musculus genomic clone RPCI-24-315L22
DNA sequence.
ACCESSION BH097469
VERSION BH097469
KEYWORDS GSS.
SOURCE GSS.
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 655)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P., and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-315L22.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 315 row: L column: 22
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..655
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-315L22"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; site_1: BamHI; site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 180 a 157 c 148 g 170 t
ORIGIN

Query Match 94.4%; Score 17; DB 13; Length 655;
Best Local Similarity 100.0%; Pred. No. 2.4e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;
QY 1 ccacagtcgtacaccac 17
|||||
Db 116 CCACAGTCGTACACCAC 132
RESULT 5
AW346676/c
LOCUS
DEFINITION 29231 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW346676
VERSION AW346676.1 GI:6844386
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 155)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAG
Plate: 13 row: B column: 12
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..155
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DHI0B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 34 a 45 c 39 g 37 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 10; Length 155;
Best Local Similarity 94.4%; Pred. No. 3.5e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1;
QY 1 ccacagtcgtacaccac 18
|||||
Db 112 CCACAGTCGTACGCCACG 95
RESULT 6
AW346655/c
LOCUS
DEFINITION 29171 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW346655
VERSION AW346655.1 GI:6844365
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 166)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAG
Plate: 13 row: H column: 6
Seq primer: ATTTAGGTGACACTATAG.

TITLE
JOURNAL
COMMENT
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAG
Plate: 13 row: H column: 6
Seq primer: ATTTAGGTGACACTATAG.

```

FEATURES
  source
    Location/Qualifiers
      1..166
        /organism="Sus scrofa"
        /db_xref="taxon:9823"
        /clone_lib="MARC 1P1C"
        /tissue_type="pooled"
        /lab_host="DH10B"
        /note="vector: pCMV SPORT6; site_1: XbaI; site_2: XhoI;
        library made from pooled tissues from day 11, 13, 15, 20,
        and 30 embryos."
      45 a 45 c 39 g 37 t
      BASE COUNT
      ORIGIN
        1..232
          /organism="Mus musculus"
          /strain="C57BL/6J"
          /db_xref="taxon:10090"
          /clone_lib="RIKEN full-length enriched, adult male thymus"
          /sex="male"
          /tissue_type="thymus"
          /dev_stage="adult"
          /lab_host="DH10B"
          /note="Site_1: SalI; Site_2: BamHI; cDNA library was
          prepared and sequenced in Mouse Genome Encyclopedia
          Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in
          RIKEN. Division of Experimental Animal Research in Riken
          contributed to prepare mouse tissues. 1st strand cDNA was
          primed with a primer [5'
          GAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
          prepared by using trehalose thermo-activated reverse
          transcriptase and subsequently enriched for full-length by
          cap-trapper. cDNA went through one round of normalization
          to Rot = 10.0 and subtraction to Rot = 100.0. Second
          strand cDNA was prepared with the primer adapter of
          sequence [5' GAGAGAGATTCGAGTTAATTAATATATCCCCCCCCC
          3']. cDNA was cloned into the XhoI and BamHI sites.
          Vector: a modified pBluescript KS(+) after bulk excision
          from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
          BamHI."
      49 a 64 c 39 g 80 t
      BASE COUNT
      ORIGIN
        Query Match 91.1%; Score 16.4; DB 10; Length 232;
        Best Local Similarity 94.4%; Pred. No. 3.8e+02;
        Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
        QY 1 ccacagtcgtacaccacg 18
        Db 190 CCACATTCGTACACCACG 173
        RESULT 8
        AV289929/c 254 bp mRNA EST 09-NOV-1999
        LOCUS AV289929 RIKEN full-length enriched, 11 days pregnant adult female
        ovary and uterus Mus musculus cDNA clone 5033430G19, mRNA sequence.
        DEFINITION AV289929
        VERSION AV289929
        KEYWORDS EST
        SOURCE AV289929.1 GI:6303960
        ORGANISM house mouse.
        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
        REFERENCE 1 (bases 1 to 254)
        AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
        Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
        Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
        ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
        Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
        Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
        ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
        Suzuki,H., Takahashi,F., Tateo,M., Tomimaga,N., Tsunoda,Y.,
        Watahiki,A., Watanabe,S., Yano,M., Yano,M., Yano,M., Yano,M.,
        Yoshiki,A., Yoshino,M., Yamamura,T., Yasunishi,A., Yokota,T.,
        YOSHIKI,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
        RIKEN Mouse ESTs (Konno,H., et al. 1999)
        Unpublished (1999)
        Contact: Yoshihide Hayashizaki
        Laboratory for Genome Exploration Research Group, RIKEN Genomic
        Sciences Center(GSC), Yokohama Institute
        The Institute of Physical and Chemical Research (RIKEN)
        1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
        Tel: 81-45-503-9222
        Fax: 81-45-503-9216
        Email: genome-res@gsr.riken.go.jp,
        URL: http://genome.gsc.riken.go.jp/
        Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
        Watsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
        ,Y.
        Transcriptional sequencing: A method for DNA sequencing using RNA
        polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
        Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
        Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
        ,Y. and Hayashizaki,Y.
        Automated filtration-based high-throughput plasmid preparation
        system. Genome Res. 9 (5), 463-470 (1999)
        Carninci,P. and Hayashizaki,Y.
        High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
        19-44 (1999)
        Please visit our web site (http://genome.rtc.riken.go.jp) for
        further details.
        Location/Qualifiers
  
```

FEATURES

Location/Qualifiers

Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
 Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki
 Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

FEATURES

Location/Qualifiers
 1. .254
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="5033430G19"
 /clone_lib="RIKEN full-length enriched, 11 days pregnant
 adult female ovary and uterus"
 /sex="female"
 /tissue_type="ovary and uterus"
 /dev_stage="11 days pregnant, adult"
 /lab_host="DH10B"
 /note="Site_1: Sali; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 100.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGATTCGAGTTAATAATATCCCCCCCCCCC
 3']. cDNA was cloned into the XhoI and BamHI sites.
 Vector: a modified pBluescript KS(+) after bulk excision
 from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
 BamHI."

BASE COUNT 51 a 74 c 45 g 83 t
 ORIGIN

Query Match 91.1%; Score 16.4; DB 10; Length 254;
 Best Local Similarity 94.4%; Pred. No. 3.9e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 ccacagtcgtacacacg 18
 ||||||| |||||||
 Db 212 CCACAGTCATACACACG 195

RESULT 9
 B1184980
 LOCUS 261 bp mRNA EST 10-JUL-2001
 DEFINITION B1184980 UNL-P-FN-dc-h-12-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 UNL-P-FN-dc-h-12-0-UNL 3', mRNA sequence.
 ACCESSION B1184980
 VERSION B1184980.1 GI:14659389
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 261)

AUTHORS

Caetano, A.R., Johnson, R.K. and Pomp, D.
 Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 Unpublished (2001)
 Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 normalized porcine ovarian follicles library. The following
 repetitive elements were found in this cDNA sequence: 255-304,
 >AT-rich#Low_complexity
 Seq primer: M13 -29
 POLYA=Yes.

FEATURES

Location/Qualifiers
 1. .261
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection
 Lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-dc-h-12-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
 library is a normalized library representing porcine
 ovarian follicles, ranging between 2.0 to 10.0 mm in
 diameter, collected during 7 days of the follicular phase
 of the pig estrous cycle. This library was derived from
 the library UNL-P-F2. The tag is a string of 5-6
 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996.
 TAG_L1B=UNL-P-FN
 TAG_TISSUE=porcine ovarian follicles
 TAG_SEQ=CACACT"

BASE COUNT 64 a 52 c 66 g 78 t 1 others
 ORIGIN

Query Match 91.1%; Score 16.4; DB 11; Length 261;
 Best Local Similarity 94.4%; Pred. No. 3.9e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 ccacagtcgtacacacg 18
 ||||||| |||||||
 Db 64 CCACAGTCATACACACG 81

RESULT 10
 AZ113096
 LOCUS 290 bp DNA GSS 12-MAY-2000
 DEFINITION RPCI-23-475F22.TJ RPCI-23 Mus musculus genomic clone RPCI-23-475F22
 DNA sequence.
 ACCESSION AZ113096
 VERSION AZ113096.1 GI:7772367
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 290)
 REFERENCE Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret
 , B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-475F22.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 475 row: F column: 22
 Seq primer: SP6
 Class: BAC ends.

FEATURES

source
 Location/Qualifiers
 1. .290
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="RPCI-23-475F22"
 /clone="RPCI-23-475F22"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1; EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 84 a 59 c 84 g 63 t
 ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 290;
 Best Local Similarity 94.4%; Pred. No. 4e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 ccacagtcgtacacacg 18
 ||||| |||||
 Db 191 CCACAGTCATACACG 208

RESULT 11
 BF713588
 LOCUS 294 bp mRNA EST 02-JAN-2001
 DEFINITION MI-P-02-axd-e-01-1-UM.s1 MI-P-02 Sus scrofa cDNA clone
 MI-P-02-axd-e-01-1-UM 3', mRNA sequence.
 ACCESSION BF713588
 VERSION BF713588.1 GI:12013063
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 294)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tugle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401

Email: cktugle@iastate.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source
 Location/Qualifiers
 1. .294
 /organism="Sus scrofa"
 /strain="crossbreed"
 /db_xref="taxon:9823"
 /clone="MI-P-02-axd-e-01-1-UM"
 /clone_lib="MI-P-02"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1; Not I; Site: 2; EcoRI; The MI-P-02 library is derived from ovary at estrus day 5. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.
 TAG_SEQ=None found"

BASE COUNT 69 a 61 c 77 g 87 t
 ORIGIN

Query Match 91.1%; Score 16.4; DB 11; Length 294;
 Best Local Similarity 94.4%; Pred. No. 4e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccacagtcgtacacacg 18
 ||||| |||||
 Db 62 CCACAGTCGTACGCCAG 79

RESULT 12
 BI184279

LOCUS 300 bp mRNA EST 10-JUL-2001
 DEFINITION UNL-P-FN-CK-g-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 UNL-P-FN-CK-g-02-0-UNL 3', mRNA sequence.
 ACCESSION BI184279
 VERSION BI184279.1 GI:14658688
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 300)
 AUTHORS Caetano,A.R., Johnson,R.K. and Pomp,D.
 TITLE Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail.
 Seq primer: M13 -29
 POLYA=Yes.

FEATURES

source
 Location/Qualifiers
 1. .300
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection
 Lines"

/db_xref="taxon:9823"
/clone="UNL-P-FN-ck-g-02-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F2. The tag is a string of 5-6
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996.

BASE COUNT 69 a 62 c 71 g 97 t 1 others
ORIGIN TAG_SEQ=None found"

Query Match 91.1%; Score 16.4; DB 11; Length 300;
Best Local Similarity 94.4%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccacg 18
||||| |||||||
DB 59 CCACAGTCATACACCACG 76

RESULT 13
AZ313323/c
LOCUS AZ313323 DNA GSS 29-SEP-2000
DEFINITION IM0029A06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0029A06 R, DNA sequence.

ACCESSION AZ313323
VERSION AZ313323.1 GI:10358137
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 370)
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0029 row: A column: 06
Seq primer: CACACGGAACAGCATATGACC
Class: plasmid ends
High quality sequence stop: 370.
Location/Qualifiers
1. .370
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0029A06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWB42nv; Purified genomic DNA from M.

FEATURES
source
1. .379
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0046N11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWB42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnates/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 66 a 100 c 84 g 120 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 370;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccacg 18
||||| |||||||
DB 165 CCACGGTCGTACACCACG 148

RESULT 14
AZ793176
LOCUS AZ793176 DNA GSS 16-FEB-2001
DEFINITION 2M0046N11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0046N11 F, DNA sequence.

ACCESSION AZ793176
VERSION AZ793176.1 GI:12937866
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 379)
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0046 row: N column: 11
Seq primer: CGTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 379.
Location/Qualifiers
1. .379
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0046N11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

FEATURES
source
1. .379
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0046N11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 120 a 80 c 95 g 84 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 379;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccag 18
||||||| |||||||
Db 75 CCACAGTCATACACCAG 92

RESULT 15

AZ831760 395 bp DNA GSS 20-FEB-2001
LOCUS 2M011K09R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M011K09 R, DNA sequence.

ACCESSION AZ831760
VERSION AZ831760.1 GI:13001668
KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 395)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

CONTACT: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0111 row: K column: 09

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 395.

Location/Qualifiers

1..395

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M011K09"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

FEATURES

source

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 121 a 88 c 112 g 74 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 395;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccag 18
||||||| |||||||
Db 253 CCACAGTCATACACCAG 270

Search completed: February 23, 2002, 23:37:23
Job time: 9444 sec

